DOCUMENT RESUME

ED 081 636 SE 016 779

AUTHOR	Paulson, James A.
TITLE	An Evaluation of Instructional Strategies in a Simple Learning Situation.
INSTITUTION	Stanford Univ., Calif. Inst. for Mathematical Studies in Social Science.
SPONS AGENCY	Office of Naval Research, Washington, D.C. Personnel and Training Research Programs Office.
REPORT NO	NR-154-326; TR-209
PUB DATE	30 Jul 73
NOTE	93p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Educational Psychology; *Instruction; *Learning; Learning Theories; *Mathematical Models; *Research; Teaching Methods
IDENTIFIERS	Research Reports

ABSTRACT

Three different strategies of choosing items for presentation in a simple list learning situation are compared. The instructional task was to teach the correct response to a number of stimulus items, using a paired-associate teaching procedure. Only one item could be presented on a given trial and the total number of trials was limited. The optimization problem considered was to find the best strategy for deciding which item to present a subject on a given trial, based on his performance on previous trials. The optimization criterion was the number of items retained on a posttest. Three progressively more sophisticated models of the learning process, the linear model, the one-element model, and a forgetting model, suggested different candidates for the optimal procedure. This paper is devoted to the derivation of theoretical expressions for operating characteristics of interest for each of the three strategies. These expressions, based on the most adequate model (the forgetting model), then permit direct numerical comparisons of the predicted performance of the three strategies for specified values of the model parameter. If the material to be learned was relatively easy, then the theoretical differences between the strategies were relatively insignificant. If the material to be learned was difficult, then the theoretical differences between the strategies were quite significant. The differences depended on the parameter values of the forgetting model in a systematic way. (Author/DT)

ED 081636

AN EVALUATION OF INSTRUCTIONAL STRATEGIES IN A SIMPLE LEARNING SITUATION

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

BY

JAMES A. PAULSON

TECHNICAL REPORT NO. 209

JULY 30, 1973

PSYCHOLOGY AND EDUCATION SERIES

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

SE 016 779

FILMED FROM BEST AVAILABLE COPY

TECHNICAL REPORTS

PSYCHOLOGY SERIES

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES

(Place of publication shown in parentheses; if published title is different from title of Technical Report, this is also shown in parentheses.)

(For reports no. 1 - 44, see Technical Report no. 125.)

50 R. C. Atkinson and R. C. Calfee. Mathematical learning theory. January 2, 1963. (In B. B. Wolman (Ed.), Scientific Psychology. New York: Basic Books, Inc., 1965. Pp. 254-275)

51 P. Suppes, E. Crothers, and R. Weir. Application of mathematical learning theory and linguistic analysis to vowel phoneme matching in Russian words. December 28, 1962.

52 R. C. Atkinson, R. Calfee, G. Sommer, W. Jeffrey and R. Shoemaker. A test of three models for stimulus compounding with children. January 29, 1963. (J. exp. Psychol., 1964, 67, 52-58)

53 E. Crothers. General Markov models for learning with inter-trial forgetting. April 8, 1963.

54 J. L. Myers and R. C. Atkinson. Choice behavior and reward structure. May 24, 1963. (Journal math. Psychol., 1964, 1, 170-203)

55 R. E. Robinson. A set-theoretical approach to empirical meaningfulness of measurement statements. June 10, 1963.

56 E. Crothers, R. Weir and P. Palmer. The role of transcription in the learning of the orthographic representations of Russian sounds. June 17, 1963.

57 P. Suppes. Problems of optimization in learning a list of simple items. July 22, 1963. (In Maynard W. Shelly, II and Glenn L. Bryan (Eds.), Human Judgments and Optimality. New York: Wiley. 1964. Pp. 116-126)

58 R. C. Atkinson and E. J. Crothers. Theoretical note: all-or-none learning and intertrial forgetting. July 24, 1963.

59 R. C. Calfee. Long-term behavior of rats under probabilistic reinforcement schedules. October 1, 1963.

60 R. C. Atkinson and E. J. Crothers. Tests of acquisition and retention, axioms for paired-associate learning. October 25, 1963. (A comparison of paired-associate learning models having different acquisition and retention axioms, J. math. Psychol., 1964, 1, 285-315)

61 W. J. McGill and J. Gibbon. The general-gamma distribution and reaction times. November 20, 1963. (J. math. Psychol., 1965, 2, 1-18)

62 M. F. Norman. Incremental learning on random trials. December 9, 1963. (J. math. Psychol., 1964, 1, 336-351)

63 P. Suppes. The development of mathematical concepts in children. February 25, 1964. (On the behavioral foundations of mathematical concepts. Monographs of the Society for Research in Child Development, 1965, 30, 60-96)

64 P. Suppes. Mathematical concept formation in children. April 10, 1964. (Amer. Psychologist, 1966, 21, 139-150)

65 R. C. Calfee, R. C. Atkinson, and T. Shelton, Jr. Mathematical models for verbal learning. August 21, 1964. (In N. Wiener and J. P. Schoda (Eds.), Cybernetics of the Nervous System: Progress in Brain Research. Amsterdam, The Netherlands: Elsevier Publishing Co., 1965. Pp. 333-349)

66 L. Keller, M. Cole, C. J. Burke, and W. K. Estes. Paired associate learning with differential rewards. August 20, 1964. (Reward and information values of trial outcomes in paired associate learning. (Psychol. Monogr., 1965, 79, 1-21)

67 M. F. Norman. A probabilistic model for free-responding. December 14, 1964.

68 W. K. Estes and H. A. Taylor. Visual detection in relation to display size and redundancy of critical elements. January 25, 1965, Revised 7-1-65. (Perception and Psychophysics, 1966, 1, 9-16)

69 P. Suppes and J. Donio. Foundations of stimulus-sampling theory for continuous-time processes. February 9, 1965. (J. math. Psychol., 1967, 4, 202-225)

70 R. C. Atkinson and R. A. Kinchla. A learning model for forced-choice detection experiments. February 10, 1965. (Br. J. math stat. Psychol., 1965, 18, 184-206)

71 E. J. Crothers. Presentation orders for items from different categories. March 10, 1965.

72 P. Suppes, G. Groen, and M. Schlag-Rey. Some models for response latency in paired-associates learning. May 5, 1965. (J. math. Psychol., 1966, 3, 99-128)

73 M. V. Levine. The generalization function in the probability learning experiment. June 3, 1965.

74 D. Hansen and T. S. Rodgers. An exploration of psycholinguistic units in initial reading. July 6, 1965.

75 B. C. Arnold. A correlated urn-scheme for a continuum of responses. July 20, 1965.

76 C. Izawa and W. K. Estes. Reinforcement-test sequences in paired-associate learning. August 1, 1965. (Psychol. Reports, 1966, 18, 879-919)

77 S. L. Blehart. Pattern discrimination learning with Rhesus monkeys. September 1, 1965. (Psychol. Reports, 1966, 19, 311-324)

78 J. L. Phillips and R. C. Atkinson. The effects of display size on short-term memory. August 31, 1965.

79 R. C. Atkinson and R. M. Shiffrin. Mathematical models for memory and learning. September 20, 1965.

80 P. Suppes. The psychological foundations of mathematics. October 25, 1965. (Colloques Internationaux du Centre National de la Recherche Scientifique. Editions du Centre National de la Recherche Scientifique, Paris: 1967. Pp. 213-242)

81 P. Suppes. Computer-assisted instruction in the schools: potentialities, problems, prospects. October 29, 1965.

82 R. A. Kinchla, J. Townsend, J. Vellott, Jr., and R. C. Atkinson. Influence of correlated visual cues on auditory signal detection. November 2, 1965. (Perception and Psychophysics, 1966, 1, 67-73)

83 P. Suppes, M. Jerman, and G. Groen. Arithmetic drills and review on a computer-based teletype. November 5, 1965. (Arithmetic Teacher, April 1966, 303-309.

84 P. Suppes and L. Hyman. Concept learning with non-verbal geometrical stimuli. November 15, 1965.

85 P. Holland. A variation on the minimum chi-square test. (J. math. Psychol., 1967, 3, 377-413).

86 P. Suppes. Accelerated program in elementary-school mathematics -- the second year. November 22, 1965. (Psychology in the Schools, 1966, 3, 294-307)

87 P. Lorenzen and F. Binford. Logic as a dialogical game. November 29, 1965.

88 L. Keller, W. J. Thomson, J. R. Tweedy, and R. C. Atkinson. The effects of reinforcement interval on the acquisition of paired-associate responses. December 10, 1965. (J. exp. Psychol., 1967, 73, 268-277)

89 J. I. Yellott, Jr. Some effects on noncontingent success in human probability learning. December 15, 1965.

90 P. Suppes and G. Groen. Some counting models for first-grade performance data on simple addition facts. January 14, 1966. (In J. M. Scandura (Ed.), Research in Mathematics Education. Washington, D. C.: NCTM, 1967. Pp. 35-43.

91 P. Suppes. Information processing and choice behavior. January 31, 1966.

92 G. Groen and R. C. Atkinson. Models for optimizing the learning process. February 11, 1966. (Psychol. Bulletin, 1966, 66, 309-320)

93 R. C. Atkinson and D. Hansen. Computer-assisted instruction in initial reading: Stanford project. March 17, 1966. (Reading Research Quarterly, 1966, 2, 5-25)

94 P. Suppes. Probabilistic inference and the concept of total evidence. March 23, 1966. (In J. Hintikka and P. Suppes (Eds.), Aspects of Inductive Logic. Amsterdam: North-Holland Publishing Co., 1966. Pp. 49-65.

95 P. Suppes. The axiomatic method in high-school mathematics. April 12, 1966. (The Role of Axiomatics and Problem Solving in Mathematics. The Conference Board of the Mathematical Sciences, Washington, D. C. Ginn and Co., 1966. Pp. 69-76.

(Continued on inside back cover)

ED 081636

AN EVALUATION OF INSTRUCTIONAL STRATEGIES IN A SIMPLE LEARNING SITUATION

by

James A. Paulson

TECHNICAL REPORT NO. 209

July 30, 1973

PSYCHOLOGY & EDUCATION SERIES

Reproduction in Whole or in Part is Permitted for Any Purpose of the United States Government

This research was sponsored by the Personnel and Training Research Programs, Psychological Sciences Division, Office of Naval Research, under Contract No. N00014-67-A-0112-0054, Contract Authority Identification Number, NR No. 154-326.

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)

Institute for Mathematical Studies in the Social Sciences - Stanford University
Stanford, California 94305

2a. REPORT SECURITY CLASSIFICATION

Unclassified

2b. GROUP

3. REPORT TITLE

An Evaluation of Instructional Strategies in a Simple Learning Situation

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report

5. AUTHOR(S) (First name, middle initial, last name)

James A. Paulson

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
30 July 1973	81	18

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
N00014-67-A-0012-0054	Technical Report No. 209
b. PROJECT NO. NR 154-326	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Personnel and Training Research Programs, Office of Naval Research Arlington, VA 22217

13. ABSTRACT

This paper compares three different strategies of choosing items for presentation in a simple list learning situation. The instructional task is to teach the correct response to a number of stimulus items, using a paired-associate teaching procedure. Only one item can be presented on a given trial and the total number of trials is limited. The optimization problem considered is to find the best strategy for deciding which item to present a subject on a given trial, based on his performance on previous trials. The optimization criterion is the number of items retained on a posttest. Three progressively more sophisticated models of the learning process, the linear model, the one-element model, and a forgetting model, suggest different candidates for the optimal procedure. The paper is devoted to the derivation of theoretical expressions for operating characteristics of interest for each of the three strategies. These expressions, based on the most adequate model, the forgetting model, then permit direct numerical comparisons of the predicted performance of the three strategies for specified values of the model parameter. If the material to be learned is relatively easy, then the theoretical differences between the strategies are relatively insignificant. If the material to be learned is difficult, then the theoretical differences between the strategies are quite significant. The differences depend on the parameter values of the forgetting model in a systematic way.

DD FORM 1 NOV 65 1473 (PAGE 1)

0102-014-6600

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Instructional Strategies Optimization of Instruction Mathematical Models of Learning Computer-assisted Instruction (CAI)						

CHAPTER I

INTRODUCTION

One obvious aim of educational psychology is to seek optimal teaching strategies for certain recurring instructional situations, based on knowledge of the learning process involved. Undoubtedly, this aim is implicit in most of the experimental work in this area. It is only recently, however, that there have been serious efforts at formal derivation of teaching strategies from descriptive models of learning processes. There are a number of good reasons why formal study of this problem has been neglected. Before formal derivation of a strategy can begin, an explicit, descriptively adequate model of the learning process under consideration must exist. Such models have been developed only in the last twenty years for even the simplest learning situations. Given an adequate descriptive framework, it is still necessary to formulate the optimization problem in terms amenable to mathematical analysis. The developments in sequential decision theory and mathematical programming (which now make such analyses feasible) have all occurred since 1945. Finally, formal optimization questions were completely academic prior to the development of modern computer technology. The large amount of record keeping and simple calculation which must be accomplished in brief time periods in order to implement optimal procedures effectively limits the use of these procedures to computer-assisted instruction settings. Now that computer-assisted instruction is becoming more widespread, optimization questions assume practical importance. This paper

is intended as a contribution to the study of an interesting optimization problem in a learning task which commonly occurs in instruction.

If appropriate mathematical tools are to be brought to bear on an optimization problem, it is necessary to place some rather severe restrictions on the nature of the learning situation to be considered. The present study is limited to situations in which the task is to teach the correct responses to a number of stimulus items, using a paired-associate teaching procedure. It is assumed that the items are learned independently, in the sense that the difficulty of learning an unknown item does not depend on whether or not other items are known. Only one item can be presented on a given trial and the total number of trials is limited. The optimization problem to be considered is to find the best strategy for deciding which item to present a subject on a given trial, based on his performance on previous trials.

There are two principal reasons for concentrating on the item-selection problem for paired-associate learning instead of considering other learning paradigms which can be described reasonably well by existing models, e.g., simple cue learning. First, the optimization problem for the paired-associate case has received a fair amount of both theoretical and empirical attention and the direction in which more study is needed is fairly clear. Second, this paradigm is directly relevant to some practical learning tasks, such as drill activities used in the learning of vocabulary items in second-language learning, the acquisition of a sight vocabulary and a knowledge of phonics in initial reading, and the mastery of spelling.

Two Strategies from Two Models

In subsequent chapters, three different strategies for choosing items to present will be examined in some detail. The first two strategies are based directly on corresponding simple models of the learning process. The third strategy is also motivated by model considerations, but the connection between model and strategy is not as direct in this case.

The first strategy may be described as follows. On a given trial, present the item which has received the fewest presentations up to that point. If more than one item satisfies this criterion, select the item at random from the set satisfying the criterion. Upon examination, this strategy is seen to be equivalent to the standard cyclic presentation procedure commonly employed in experiments on paired-associate learning. It amounts to presenting all items once, randomly reordering them, presenting them again and repeating the procedure until the number of trials allocated to instruction have been exhausted. This procedure will be referred to hereafter as the RC (for random cyclic) procedure or strategy.

A rationale for the RC procedure can be provided by a simple linear model of the learning process. In this model, the state of the learner with respect to each item in a list depends only on the number of times each item has been presented. The state of the learner is represented by his momentary probability of error for each item. At the start of instruction, all items have some initial probability of error, say q_1; each time an item is presented, its error probability is reduced by a factor α, which is less than one. That is,

$$q_{n+1} = \alpha q_n$$

or alternatively

$$q_{n+1} = \alpha^n q_1 .$$

When an item is presented for the n^{th} time, the reduction in error probability is given by

$$q_n - q_{n+1} = \alpha^{n-1}(1-\alpha)q_1 .$$

The fact that the decrement in error probability for an item becomes smaller each time it is presented leads naturally to the RC procedure.

The second strategy is more complicated to describe, but the essential idea is very simple: ignore responses prior to the last error on an item; present the item which has received the fewest correct responses since its last error. If more than one item is eligible according to this rule, select the item to be presented at random from the set of eligible items. Karush and Dear (1966) proved that this strategy is optimal if the assumption that learning proceeds according to the so-called one-element model is valid. The strategy is optimal in the sense that it maximizes the expected number of items learned in a fixed number of presentations. For this reason, this strategy will be referred to hereafter as the OEM strategy.

According to the one-element model a student is in one of two states with respect to each item at any given point in time: the learned state or the unlearned state. When an unlearned item is presented, it moves into the learned state with probability c. That is,

$$q_{n+1} = \begin{cases} q_n , & \text{with probability } 1-c \\ 0 , & \text{with probability } c . \end{cases}$$

Once an item is learned, it remains in the learned state throughout the course of instruction, so there is no reason to present the item again. A subject may respond correctly to an item even though the item is in the unlearned state and the subject is guessing. In effect, the OEM strategy selects items for presentation that are most likely to be in the unlearned state.

The linear model, which provides a basis for the RC procedure, and the one-element model, which provides a basis for the OEM procedure, are the simplest models for paired-associate learning having any empirical support. On the whole, the one-element model gives a better account of data from experiments using the RC procedure, than the linear model; see, for example, Bower (1961). The data which lead to this conclusion would also lead one to believe that a given number of presentations allocated to a list of items using the OEM procedure would produce significantly better results than the same number of presentations allocated according to the RC procedure. The predicted advantage for the OEM strategy often fails to materialize, unless special modifications are made in the OEM procedure. This anomaly provides the motivation for the developments to be reported in subsequent chapters.

Atkinson and Crothers (1964) reported data comparing performance of several models of learning and retention which suggests consideration should be given to procedures based on models taking forgetting phenomena into account. However, it turns out that the performance of procedures

based directly on forgetting models is difficult to characterize in a general way. The two strategies described above are special in two respects, which make the relationship between model and strategy simpler for them than it is in general. One special feature both the OEM and RC procedure possess is that they maximize both immediate gain in probability of correct response and global gain over the course of the experiment considered as a whole. It is the exception rather than the rule for a procedure to be capable of maximizing both of these quantities. Another rather rare property which these procedures have in common is that implementation of neither one depends on the parameter values of the model on which it is based.

The approach to be taken in this paper is to use a general theory of learning and forgetting to describe performance of three strategies, the two already mentioned and a third hypothetical strategy. This hypothetical strategy is a modification of the OEM strategy which would be optimal under the assumption of the general forgetting model if it could be carried out. For reasons which will be clear when the strategy is described in detail later, the strategy would be impossible to implement. Nevertheless, the strategy serves as a useful bound against which implementable strategies can be compared.

The rest of this paper is devoted to consideration of the relative performance of the three strategies discussed above, using the general forgetting theory as a model framework. The forgetting theory and the reasons for adopting it for this study are described in Chapter II; this chapter also contains several counterexamples which demonstrate the infeasibility of dealing with globally optimal strategies in the context

of the forgetting theory framework. Because it is not feasible to deal with globally optimal strategies, it is necessary to take a descriptive approach to the evaluation of presentation strategies. Chapter III is devoted to the definition and derivation of formulas for operating characteristics of the three strategies described above. These formulas are then used in Chapter IV to make numerical comparisons of the strategies for selected special cases within the general framework of forgetting theory. In Chapter V the conclusions of the study are summarized and implications for future research are considered. The results presented in this paper are new, at least to the author's knowledge, unless otherwise noted.

CHAPTER II

A GENERAL FORGETTING THEORY AND ITS IMPLICATIONS FOR PRESENTATION STRATEGIES

The General Forgetting Theory proposed by Rumelhart (hereafter to be called the GFT) is a synthesis of several models which represent different ways to generalize the OEM. These models retain much of the simple all-or-none character of the OEM while introducing the factor of forgetting, which the OEM does not take into account. The theory is relevant to the problem being treated here because the phenomena of forgetting could work in a variety of ways to undermine the OEM strategy.

There is considerable evidence now that forgetting works in such a way that the OEM strategy is not as effective as it theoretically should be. In an experiment designed to test the advantage of the OEM procedure over the standard RC procedure reported by Dear, Silberman, Estavan and Atkinson (1967), the advantage predicted for the OEM was not observed. Experimental findings of Hellyer (1962) and Greeno (1964), among others, suggest that when items are presented repeatedly within a short period of time (as they often are under the OEM procedure), many are responded to correctly on the massed presentations, but are then rapidly forgotten. This interpretation is consistent with that of Dear and his associates. Experiments which have shown the predicted advantage for the OEM procedure, such as one by Lorton reported in Atkinson and Paulson (1972), have modified the procedure to minimize the number of massed presentations.

The General Forgetting Theory

The GFT can be described as follows. At any given time, a subject is in one of three possible states of learning with respect to each item:

the unlearned state, the short-term retention state, or the long-term retention state. When an item is presented, transitions between states occur according to the following stochastic matrix.

		State on trial t+1			Probability of correct response, given the state
		L	S	U	
State on trial t	L	1	0	0	1
	S	c	1-c	0	1
	U	a	b	1-a-b	g

That is to say, if an unlearned item is presented, then with probability a it is learned in such a way that it will be retained for a relatively long time, with probability b it is learned in such a way that it is likely to be forgotten soon, and with probability 1-a-b it remains unlearned. If an item in the short-term retention state is presented, then with probability c it will shift to the long-term retention state, and with probability 1-c it will remain in the short-term state. Whenever an item reaches the learned state, it remains there for the duration of the experiment.

When an item in either the long- or short-term state is presented, the correct response is given with probability 1. If an unlearned item is presented, the correct response is given with the guessing probability g.

In this extended model it is necessary to consider what happens to items which are not presented on a trial. Transitions between states occur according to the matrix

$$
\begin{array}{c} \\ \text{State on trial } t \end{array}
\begin{array}{c}
\begin{array}{ccc} \text{State on trial } t+1 & & \\ \text{L} & \text{S} & \text{U} \end{array} \\
\begin{array}{c} \text{L} \\ \text{S} \\ \text{U} \end{array}
\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1-f & f \\ 0 & 0 & 1 \end{bmatrix}
\end{array}
$$

That is, items in the short-term retention state are forgotten, with probability f, while items in the long-term retention state or the unlearned state are unaffected.

If it is stipulated that the parameter b in the learning matrix is 0, then an item can never enter the short-term state so the model reduces to the OEM in this case.

Perhaps the quickest way to follow the character of the GFT framework is to consider briefly some of the other models which it encompasses as special cases. Table 2.1 is intended to give the reader an abbreviated natural history of the formulation just given. The various special cases all assume a forgetting matrix of the form given above. They differ with respect to the form of the learning matrix. The differences reflect differences in assumptions regarding two separate issues.

The earlier models all assume that some learning takes place whenever an item in the unlearned state is presented. Hence, the probability of staying in the unlearned state is 0. The models differ regarding the relative size of the probability of transition to the long-term state from the unlearned and short-term states, respectively. The model of Atkinson and Crothers (1964) assumes that these transition probabilities are equal. The model of Greeno (1964) assumes that transitions to the

Table 2.1

Transition Matrices of the Learning Process for Special Cases of the General Forgetting Theory

		State on trial N+1			Comments
		L	S	U	
State on trial N	L	1	0	0	One of the original versions of the long-short model due to Atkinson & Crothers (1964)
	S	a	1-a	0	
	U	a	1-a	0	
	L	1	0	0	A coding model due to Greeno (1964)
	S	0	1	0	
	U	a	1-a	0	
	L	1	0	0	A partial learning model due to Bernbach (1965)
	S	a	1-a	0	
	U	0	1	0	
	L	1	0	0	A general model encompassing the three above
	S	b	1-b	0	
	U	a	1-a	0	
	L	1	0	0	A further extension introducing an attention parameter γ, due to Rumelhart (1967)
	S	$b\gamma$	$1-b\gamma$	0	
	U	$a\gamma$	$(1-a)\gamma$	$1-\gamma$	
	L	1	0	0	The formulation given in this paper
	S	c	1-c	0	
	U	a	b	1-a-b	

long-term state can only take place from the unlearned state, and the model of Greenbach (1965) assumes that these transitions can only occur from the short-term state. These three models all have learning matrices depending on a single parameter. If one wishes to leave the issue of the relative size of these transition parameters open, he can do so at the price of introducing a second learning parameter, as indicated in the fourth transition matrix in Table 2.1.

In formulating his GFT, Rumelhart leaves the issue of the relative size of the transition probabilities open, and introduces a third parameter γ, which he regards as an attention parameter. If γ is less than 1, there is positive probability that an item in the unlearned state will stay there following a presentation. The final matrix, which corresponds to the one given above, is a very slight generalization of Rumelhart's formulation. Combinations of a, b, and c in the final formulation for which $c > a+b$ do not correspond to any possible combination of a, b, and γ in the fifth matrix. The cases of most concern in this paper satisfy the constraint $c \leq a$, so the difference is essentially one of notational convenience. According to Rumelhart, the introduction of γ results in a marked improvement in the fit of the model to his data.

Now let us consider the implications of the GFT framework for presentation strategies. It is well-known that the strategy which maximizes immediate gain in probability of correct response can differ from the strategy which maximizes the global gain over the course of the experiment as a whole. The latter type of strategy is called globally optimal. The rest of this chapter presents findings which together demonstrate the need to leave the search for globally optimal strategies in favor of

a detailed description of operating characteristics of certain selected strategies. The crux of the argument is that the globally optimal strategy requires looking more than one trial ahead in all cases of interest; in the context of the general forgetting theory this fact alone makes the globally optimal strategy very difficult to characterize in a useful way.

It is unfortunate that the strategy that looks just one trial ahead is not globally optimal, because this strategy is mathematically simple and intuitively reasonable. There are very clear interpretations of this strategy in each of the special cases of the GFT framework described above. Each of these is a plausible generalization of or alternative to the optimal strategy corresponding to the one-element model. But counter-examples will be provided to show that none of these are globally optimal.

Strategies Maximizing Immediate Gain

Let $\ell_{i,n}$, $s_{i,n}$, $u_{i,n}$ be the respective probabilities that item i is in the long-term, short-term, or unlearned state on trial n. Let $\delta_{i,n}$ be an indicator variable which is 1 if item i is presented on trial n, 0 if it is not. The probability that item i is in the long-term retention state on trial n+1 is given by

$$\ell_{i,n} + \delta_{i,n}(cs_{i,n} + au_{i,n}) .$$

The expected gain in number of items in the long-term state on trial n+1 is given by

$$\sum_{i=1}^{I} \ell_{i,n+1} - \sum_{i=1}^{I} \ell_{i,n} = \sum_{i=1}^{I} \delta_{i,n}(cs_{i,n} + au_{i,n}) .$$

Clearly, the expected gain is maximized if we present the items with largest values of $cs_{i,n} + au_{i,n}$.

In the special case a = c this amounts to presenting the items with the largest values of $s_{i,n} + u_{i,n}$, or the smallest value of $\ell_{i,n}$. Hence, in this case the strategy maximizing the immediate gain is a generalization of the one-element model strategy.

In the special case c = 0 the expected gain is maximized by presenting the items most likely to be in the unlearned state, which is a different generalization of the OEM strategy. In the case a = 0 immediate gain is maximized by presenting the items most likely to be in the short-term state. These comments are summarized as a theorem for future reference.

Theorem 2.1. Let ℓ_i, s_i, and u_i be the state probabilities for item i on a given trial, and let a and c be the transition probabilities for moving to the state L from state U and S, respectively. Then the expected immediate gain is maximized by presenting the items with largest values of

$$G_i = au_i + cs_i .$$

In the case a = c, this is equivalent to presenting the items least likely to be in L. In the case a = 0, it means presenting the items most likely to be in S. In the case c = 0, it means presenting the items most likely to be in U.

Counterexamples to Demonstrate Non-optimality of Maximizing Immediate Gain

When Karush and Dear (1966) established the global optimality of the OEM strategy, they did so by first deriving the strategy maximizing

the immediate gain and then showing by an induction argument that this strategy is, in fact, globally optimal. Their approach to the characterization of optimal strategies will not carry over into the GFT framework, as the counterexamples to be presented will show. A counterexample will be described for each of the special cases mentioned in Theorem 2.1. In each of these cases, the strategy maximizing immediate gain focuses all attention on a single state probability, ignoring the other two. The thrust of the counterexample in each case is to show that the other two state probabilities carry important information.

Case 1: $a = 0$. It is perhaps easiest to see the necessity for looking ahead more than one stage by examining the special case in which each item must pass through the short-term state before it can reach the long-term state. That is, the probability a of making a direct transition from the unlearned state to the long-term state equals 0. In this special case, the policy maximizing immediate gain is to present the item most likely to be in S. If all items start out in the unlearned state, all have probability zero of being in S. After the first item is presented, it has positive probability of being in S and will continue to have for the duration of the experiment. Since the other items would still have probability zero of being in S, the policy maximizing immediate gain would be to continue to present the first item indefinitely. There is no immediate gain to be had from presenting a new item once, because it will just go to the short-term state, but there may be considerable advantage in presenting it twice. The strategy maximizing immediate gain ignores this possibility.

Suppose A and B are two unknown items and four trials are available for teaching both of them. Suppose the parameter values are $a = 0$, $b = 1$, $c = f = .5$ for both items. The strategy maximizing immediate gain would devote all four trials to one item. It is easy to verify that a better strategy would be to present each item twice in succession. This example is logically sufficient to prove that the strategy maximizing immediate gain is not in general the globally optimal policy in the GFT framework. It is still possible that such strategies are globally optimal for some other special cases. Two more examples will be given to show that this is not the case in the instances of most interest in the present study.

Case 2: $a = c$. It was shown earlier that when a equals c, the strategy maximizing the immediate gain is to present the items least likely to be in L. Suppose two items, A and B, are not in L. Suppose A is in state U and B is in state S. Which item should be presented on the next trial? From the point of view of immediate gain it makes no difference, because the probability of either item making the transition to state L is the same; that is, $a = c$. However, from a longer term point of view it does make a difference. If item B is presented, it will be responded to correctly; if item A is presented, it will likely be responded to incorrectly. The incorrect response would be informative, letting the experimenter know that the item was certainly not in L before its presentation. Thus, it would be preferable to present item A.

The preceding argument is not completely satisfactory because it is assumed that both A and B are not in L, so presenting A is not as informative as it seems. But the argument would apply to a case where A and

B have the same positive probability of being in L but A is more likely to be in U than is B. This situation is likely to arise in practice. For example, consider the case where $a = b = c = f = g = .5$. The sequence of events given in Figure 2.1 consists of an initial phase and two alternative strategies for a second phase. At the end of the initial phase both items have probability about .79 of being in L, item B has probability 0 of being in U, while item A has probability .10 of being in U. The strategy maximizing immediate gain would be indifferent with regard to which of the two alternatives to follow. Direct numerical calculations demonstrate that it would be preferable to present A first.

Case 3: $c = 0$. Table 2.2 gives a sequence of events to show the need for looking more than one trial ahead in the case $c = 0$, $a = b = .25$, $f = g = .50$. (It is still a two-item list under consideration.) When $c = 0$, the strategy maximizing immediate gain is to present the items most likely to be in U. The idea behind this example is a simple one: it can be advantageous to refrain from presenting an item, even if it is the one most likely to be in U, if waiting will significantly increase the probability of being in U. It is necessary that there be another item available to present whose prospect for immediate gain is nearly as good.

This example also consists of two phases, the first phase showing how two items could come to have certain critical state probabilities under the policy of maximizing immediate gain, the second phase showing the advantage of looking two stages ahead instead of one, given these state probabilities. At the end of the initial phase the state probability vectors for items A and B are $(\frac{1}{4}, \frac{1}{4}, \frac{1}{2})$ and $(\frac{6}{16}, \frac{3}{16}, \frac{7}{16})$,

Sequence of Events

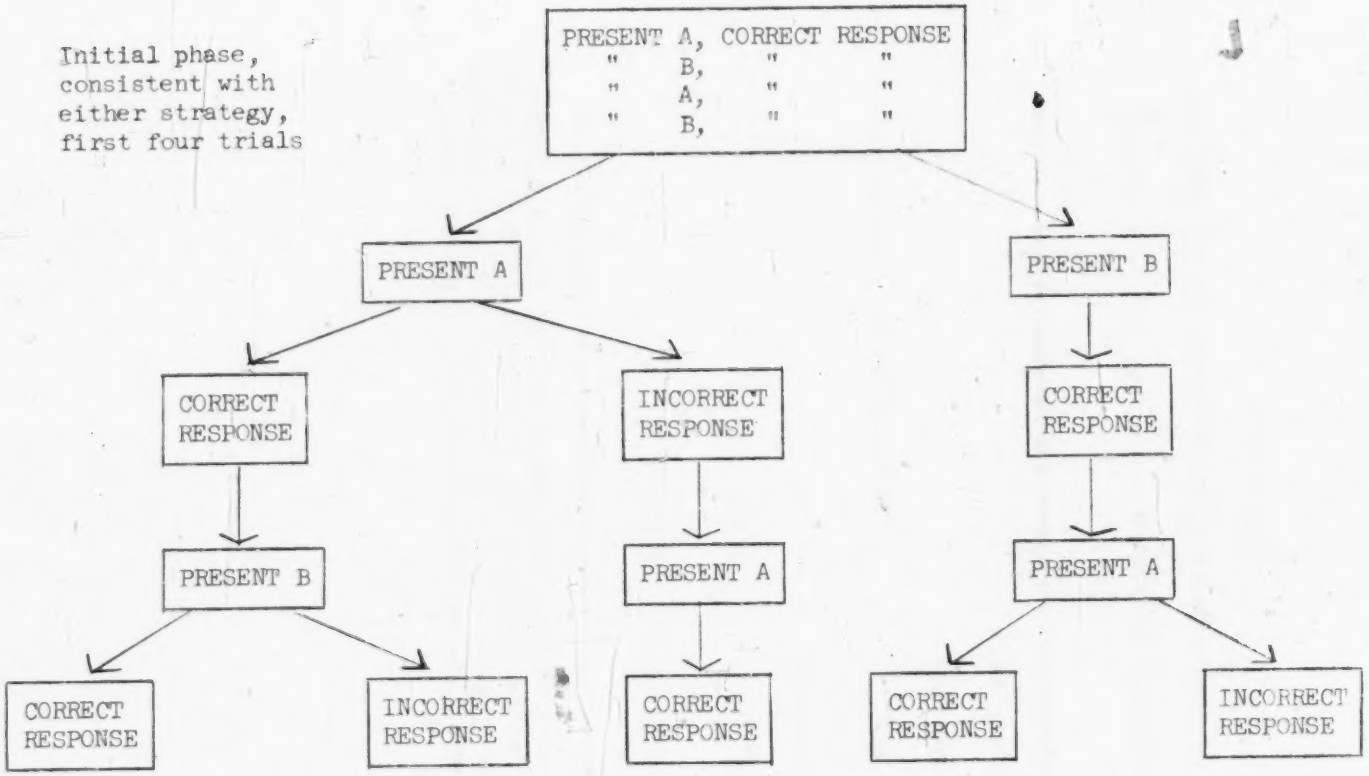

Figure 2.1. Sequence of events showing policy of maximizing immediate gain is suboptimal when a=b=c=f=g=.5.

Table 2.2

Sequences of Events Showing Policy of Maximizing Immediate Gain is Suboptimal When $a = b = .25$, $c = 0$, $f = g = \frac{1}{2}$.

Initial phase under policy maximizing immediate gain:	A+/B+/A+/B-/B+/A-/ continuation
Continuation under MIG policy:	A $\pm$/B
Better continuation:	B $\pm$/A

Note: Letter indicates item presented; the sign following the letter indicates the correctness of the response, where $\pm$ means that the correctness does not influence the decision regarding which item to present next.

respectively. Since $\frac{1}{2} > \frac{7}{16}$, the policy maximizing immediate gain would be to present A next. If A is presented, the state probabilities on the next trial will be such that item B will be presented on the final trial, whether the response to A is correct or not. Similarly, if item B is presented first, item A should be presented on the final trial. Direct calculations show that the latter policy is slightly preferable.

These examples show that globally optimal strategies in the GFT framework generally require more than maximization of immediate gain. They do <u>not</u> show that the strategy maximizing immediate gain is never optimal. It obviously is in the special case where the model reduces to the OEM, when $b = 0$. Even if b is positive, if it is sufficiently small, it will have no bearing on the optimal strategy. The implications of the counterexamples given above concern what can be said in general about globally optimal strategies without specifying the exact values of the parameters. The fact that we can say very little suggests that a descriptive approach permitting comparison of strategies with one another, but not with the globally optimal strategy, would be appropriate.

CHAPTER III

OPERATING CHARACTERISTICS OF THREE PRESENTATION STRATEGIES

A reasonable model of learning should enable one to make a variety of predictions about the overall state of a list of items, provided the items are presented in a certain way. The presentation procedure used most often in the evaluation of learning models is the RC procedure discussed earlier. When this procedure is used, sample statistics corresponding to expressions for the trial of last error, the probability of error given the last response to the item was an error, and other descriptive statistics of interest can be calculated and compared with theoretical predictions.

Matters become more complicated when presentation procedures other than the RC procedure are employed. For one thing, the meaning of the descriptive statistics which are of interest may change when other procedures, such as the OEM strategy, are used. For example, under the OEM strategy the number of presentations varies widely from one item to another, so the "trial of last error" means something different than it does under the RC strategy. Another difficulty which arises concerns the derivation of theoretical expressions for statistics of interest. Indeed, it is only in exceptional cases that it is possible to derive explicit expressions for quantities of interest. Usually, the number of times each item is presented, and when, is subject to such a variety of contingencies that explicit calculations are not feasible. As an illustration, consider the case of the strategy maximizing immediate

gain within the GPT framework. The item to be presented on a given trial is the item having the highest value on an index which is a function of the parameter values, the number of correct responses since the last error, and the number of items intervening between each of these correct responses. Direct calculation of exact theoretical formulas of interest in this situation appears to be hopeless.

The situation is not as bleak as this in the case of the OEM strategy because there is a pattern to presentations under this procedure which serves as a natural basis for summarizing the overall state of the items. This pattern will be described in some detail, because it serves as a basis for most of the theoretical derivations in this chapter.

Presentation cycles and "almost" sufficient histories. Under the OEM procedure items are presented in a series of cycles which are similar in many respects to trials under the RC procedure. Each item receives a specified treatment on each cycle. The difference between the RC and the OEM procedures lies in the fact that under the OEM procedure the treatment of an item may involve several presentations, whereas under the RC procedure treatment consists of a single presentation per item per trial.

The cyclic structure of item presentations under the OEM strategy arises in the following manner. The strategy says to present an item whenever the string of consecutive correct responses to it is shorter than the corresponding strings for the other items. If several items are tied at a given point, the choice is made on a random basis. At the beginning of a cycle this index is the same for all items. If an item is presented and receives a correct response, its index is incremented

by 1 and is therefore greater than the indices for the other items. It will not be eligible for presentation again until all the other items reach the same level, i.e., until the cycle has been completed for all the other items. If, on the other hand, the item is responded to incorrectly, its index is reset to 0, so it is lower than all other items and will continue to be lower until repeated presentations bring it back to their level.

Denote by cycle n those presentations required to move the list from the place where all items have been responded to correctly n-1 times in a row to the place where they have all been responded to correctly n times in a row. Most of the operating characteristics of interest in describing performance under the OEM procedure, such as cycle of last error, probability of error, and cumulative number of presentations, are functions of the cycle number.

When the OEM is an accurate description of the learning process, the cycle number is a sufficient history for describing the state of a list of items because for every item in the list the cycle number is equal to the number of correct responses since the last error. The lag between successive presentations of a given item is irrelevant. If forgetting is taken into account, as it is in the GFT framework being considered here, then the lags become an important factor. Strictly speaking, given a GFT model a sufficient history for each item involves the number of correct responses since the last error and the number of intervening items presented between each of these correct responses. Fortunately, it is possible to simplify this sufficient history in the case of the OEM procedure with negligible loss of information.

The following observations are the justification for the simplification of the sufficient history of an item which will be referred to as an *almost sufficient history*:

1. When an item is presented on a cycle and the response is correct on the first try, the item is not presented again on that cycle, so it is safe to assume that many intervening items will be presented before that item is presented again.
2. When an item is presented on a given cycle and the response is an error, there follows a string of presentations of the item without any intervening items, culminating in a string of presentations with correct responses whose length is 1 less than the cycle number. The last correct response is made after a number of intervening items have been presented.

As a consequence of these features of the OEM procedure, the state of an item at a given point in the instructional process is essentially determined by its *cycle number* and the *cycle of last error*. The string of correct responses on the cycle of last error has lag 0 between each presentation except for the last presentation. It and the correct responses on subsequent cycles have what may, for practical purposes, be regarded as infinite lag. Therefore, $s \approx 0$ for these presentations, where s is the probability of the item being in the short-term state. Thus, the cycle number indicates the number of consecutive responses to an item and the cycle of last error indicates the lengths of an initial block of presentations with no intervening items and a final block of presentations with many items intervening between them.

By conditioning on the cycle number and the cycle of last error, it is possible to calculate approximate theoretical expressions for a number of statistics of interest when this presentation procedure is employed. These calculations will be carried out in the next section. Subsequent sections will discuss corresponding expressions for other procedures. The other procedures to be treated include the RC procedure and the hypothetical procedure that serves as a baseline for comparisons to be made in the next chapter. This hypothetical strategy will be referred to as the modified OEM procedure. This is the procedure that would result if one were somehow able to introduce very long lags between the several presentations of an item which has been responded to incorrectly on a given cycle. This hypothetical procedure is better than any procedure that can really be carried out, so it serves as a useful bound in determining how close a suboptimal procedure is to being optimal. It serves this purpose in the place of the optimal strategy, whose operating characteristics cannot be determined in practice because the strategy itself is unknown.

Operating Characteristics of the OEM Strategy

The basic statistic for describing performance under the OEM strategy is the expected number of presentations per item required for each cycle. Every item receives exactly one presentation on the first cycle, but on subsequent cycles the number of presentations is a random variable. Define three random variables for each cycle $k \geq 2$ as follows. Let

P_k = number of presentations required for an item to complete cycle k,

W_k = number of presentations following an error required to obtain a sequence of k consecutive correct responses, and,

$\pi_{k,k}$ = the probability of at least one error on cycle k for a given item (the reason for the double subscript will become clear later).

Now

$$P_k = \begin{cases} 1 & , \text{ with probability } 1-\pi_{k,k} \\ 1+W_k, & \text{ with probability } \pi_{k,k} \end{cases}$$

Therefore, we have

(1) $$EP_k = 1 + \pi_{k,k} EW_k .$$

The key task of this section is to find $\pi_{k,k}$ and EW_k. The main ideas to be used in accomplishing this task apply to a broader class of models than the GFT framework, so they will be set forth in some generality. Then specific approximations will be obtained for the GFT framework.

General Formulas

The distribution of W_k. The crucial fact to note about W_k is that it is the waiting time in a terminating renewal process, in the sense that Feller (1969, p. 186) defines the term. A renewal process is a stochastic process whose characteristic feature is that there is an event which sets the process back to its starting point whenever it occurs. Such an event is called a recurrent event. We may regard W_k as the waiting time for the first occurrence of a sequence of k consecutive correct responses following an error. The occurrence of an error is the recurrent event which resets the probabilistic structure

of the process. Renewal theory provides a fundamental relationship between the distribution of W_k and the distribution of E_k, the waiting time for the next error. Actually, the distribution of E_k is defective because there is positive probability that the cycle will terminate before there is another error on the cycle (hence, the term terminating renewal process). For this reason, we also consider the conditional distribution of E_k, given that there is another error on the cycle.

Let

$$P(W_k = n) = w_{k,n} ,$$

and let

$$P(E_k = \nu) = e_{k,\nu} \quad \text{for} \quad \nu = 1,\ldots,k ,$$

and

$$P(\text{process terminates without an error}) = e_{k,k+1} .$$

The conditional distribution of E_k, given that another error occurs on the cycle, is given by

$$P(E_k = \nu | E_k = \nu', \text{ for some } \nu' = 1,\ldots,k) = \frac{e_{k,\nu}}{1-e_{k,k+1}} .$$

Let the conditional random variable be denoted by E_k^*. The generating function for the distribution of E_k^* is then

$$g_{E_k^*}(s) = \frac{1}{1-e_{k,k+1}} \sum_{\nu=1}^{k} e_{k,\nu} s^{\nu} .$$

Let $g_{W_k}(s) = \sum_{n=0}^{\infty} w_{k,n} s^n$ be the generating function of W_k.

Theorem 3.1. Consider any model for which the occurrence of an error on cycle k is a terminating recurrent event. Then conclusions A, B, and C below follow.

A. The distribution of W_k is given by

$$(2) \qquad w_{k,n} = \begin{cases} 0 & \text{for } n < k \\ e_{k,k+1} & \text{for } n = k \\ \sum_{\nu=1}^{k} e_{k,\nu}\, w_{k,n-\nu} & \text{for } n > k \,. \end{cases}$$

B. The generating function of W_k is given by

$$(3) \qquad g_{W_k}(s) = \frac{e_{k,k+1} s^k}{1-(1-e_{k,k+1}) g_{E_k^*}(s)} \,.$$

C. The expected value of W_k is given by

$$(4) \qquad EW_k = k + (1-e_{k,k+1}) \frac{g'_{E_k^*}(1)}{e_{k,k+1}} \,.$$

Before proceeding with the proof of the theorem, it would be good to interpret the terms on the right-hand side of Equation 4. Equation 4 states that the average number of presentations required following an error on cycle k is the sum of the number of consecutive correct responses required, k, and the number of extra responses made necessary by further errors. The latter term is the product of $1-e_{k,k+1}$, the probability of further errors on the cycle; $\frac{1}{e_{k,k+1}}$, the expected number of errors on the cycle; and $g'_{E_k^*}(1)$, the average number of presentations per error.

These quantities depend, of course, on the exact nature of the particular learning model being considered.

Proof of the theorem. It is obvious that $w_{k,n} = 0$ for $n < k$. The waiting time for k consecutive correct responses is equal to k with probability $e_{k,k+1}$. If $W_k > k$, there must have been an error on some presentation $\nu = 1,\ldots,k$. The probability that $W_k = n$, given an error on presentation ν, is $w_{k,n-\nu}$. The expression for $w_{k,n}$ when $n > k$ in Equation 2 is the weighted average of the $w_{k,n-\nu}$'s. Taken together, these comments justify Equation 2.

If both sides of Equation 2 are multiplied by s^n and the results for all values of n added together, the result is

$$
\begin{aligned}
(5) \qquad g_{W_k}(s) &= e_{k,k+1}s^k + \sum_{n=k+1}^{\infty} \sum_{\nu=1}^{k} e_{k,\nu}\, w_{k,n-\nu}\, s^n \\
&= e_{k,k+1}s^k + \sum_{\nu=1}^{k} e_{k,\nu}s^{\nu}\left(\sum_{n=k+1}^{\infty} w_{k,n-\nu}s^{n-\nu} \right) \\
&= e_{k,k+1}s^k + (1-e_{k,k+1})\, g_{E_k^*}(s)\, g_{W_k}(s) \\
&= \frac{e_{k,k+1}s^k}{1-(1-e_{k,k+1})g_{E_k^*}(s)} ,
\end{aligned}
$$

which is Equation 3.

Differentiating Equation 3 yields

$$
g'_{W_k}(s) = \frac{[1-(1-e_{k,k+1})g_{E_k^*}(s)]ke_{k,k+1}s^{k-1} + e_{k,k+1}s^k(1-e_{k,k+1})g'_{E_k^*}(s)}{[1-(1-e_{k,k+1})g_{E_k^*}(s)]^2} .
$$

Letting $s = 1$ and noting that $g_{E_k^*}(1) = 1$, we obtain

$$g'_{W_k}(1) = k + \frac{1-e_{k,k+1}}{e_{k,k+1}} g'_{E^*_k}(1) ,$$

which justifies Equation 4 and completes the proof.

Recursion formulas for computing $\pi_{k,k}$. The general formulas for computing $\pi_{k,k}$ to be given are valid for any model of learning for which the initial probability of a correct response is a guessing probability. However, they are really useful only if the model is one for which conditioning on the cycle of last error leads to a simplification or a reasonable approximation, as is the case with the GFT.

Let $q_{k,n}$ be defined as follows:

$$q_{k,n} = \begin{cases} P(\text{Error on cycle } n \mid \text{cycle } n\text{-}1 \text{ just completed, no errors on item yet}), & \text{for } k = 0, \\ P(\text{Error on cycle } n \mid \text{cycle } n\text{-}1 \text{ just completed, last error on cycle } k), & \text{for } k = 1,\ldots,n\text{-}1 . \end{cases}$$

Similarly, let $\pi_{k,n}$ be defined as

$$\pi_{k,n} = \begin{cases} P(\text{No errors on item} \mid \text{cycle } n \text{ just completed}), & \text{for } k = 0 , \\ P(\text{Last error was on cycle } k \mid \text{cycle } n \text{ just completed}), & \text{for } k = 1,\ldots,n . \end{cases}$$

Note that according to this definition $\pi_{k,k}$ is the probability that the last error was on cycle k, given that cycle k has just been completed. But this is just the probability that there was an error on cycle k, which accords with the definition of $\pi_{k,k}$ given earlier. Also, note that $q_{k,n}$ is not defined for n = 1. It is not needed and the definition makes no sense in that instance.

Theorem 3.2. $\pi_{k,k}$ can be computed in terms of $q_{j,\nu}$'s with $j,\nu < k$ using the following relationships.

(6a) $\pi_{0,1} = g \; ; \; \pi_{1,1} = 1-g \; ;$

(6b) $\pi_{0,n} = g \prod_{k=1}^{n-1} (1-q_{0,k+1}) \; , \quad \text{for} \quad n \geq 2 \; ;$

(6c) $\pi_{j,n} = \pi_{j,j} \prod_{\nu=j+1}^{n} (1-q_{j,\nu}), \quad \text{for} \quad j \geq 1, \; n > j \; ;$

(6d) $\pi_{k,k} = \sum_{i=\nu}^{k-1} \pi_{i,k-1} \, q_{i,k} \; , \quad \text{for} \quad k \geq 2 \; .$

The justification for Equation 6a is that $\pi_{0,1}$ and $\pi_{1,1}$ are the respective proportions of items having and not having correct responses on the first presentation. Since items are assumed to be unknown at the outset, the values are g and 1-g. The relationship expressed in Equation 6b says simply that the probability of no errors on an item through cycle n is the product of the probability of guessing correctly on the first cycle and the appropriate conditional probabilities of not making an error on succeeding cycles.

On completion of cycle n, where $n > j$, the proportion of items whose last error was on cycle j is the proportion of items with an error on cycle j and no further errors through the n^{th} cycle. This is the relationship expressed in Equation 6c.

The formula given in Equation 6d expresses $\pi_{k,k}$ as the sum of the conditional probabilities of error on cycle k, given the cycle of last error, each weighted by the probability that it was the last error. This completes the proof.

In order to use Theorems 1 and 2 to find the expected number of presentations on cycle k for a specific model of the learning process, it is usually necessary to have explicit expressions in terms of model parameters for the following quantities:

A. The $q_{k,n}$'s, the conditional probabilities of error on cycle n, given the last error was on cycle k.

B. The probability of no further errors on the k^{th} cycle following an error on that cycle, $e_{k,k+1}$.

C. The conditional distribution of the waiting time for the next error following an error on cycle k, given that there will be another error on the cycle, and its mean.

Before deriving expressions for these quantities in the general GFT case, it should be noted that the calculations can be simplified considerably in the important special case where Greeno's model applies. The calculation of the $q_{k,n}$'s can be avoided because a formula giving the approximation for $\pi_{k,k}$ can be derived directly. The other quantities of interest are simple functions of $\pi_{k,k}$ and the cycle number.

Theorem 3.2a. When Greeno's model applies, $\pi_{k,k}$ can be approximated by the following formula.

$$\pi_{k,k} = \begin{cases} 1-g, & \text{when } k = 1 \\ (1-g)(1-a)\left[\dfrac{g(1-a)}{a+g(1-a)}\right]^{k-2}, & \text{for } k \geq 2. \end{cases}$$

Proof. The formula is obvious for k = 1 and 2. It needs to be demonstrated that $\pi_{k+1,k+1} = \dfrac{g(1-a)}{a+g(1-a)}\,\pi_{k,k}$, for $k > 2$. It suffices to show that P (In state U on cycle k+1 | In state U on cycle k) = $\dfrac{g(1-a)}{a+g(1-a)}$.

An item must be responded to correctly in order to complete a cycle. If an item is in U at the start of a cycle, one of three things will happen: a correct response by guessing and no transition to the long-term state (with probability g(1-a)); a response, correct or incorrect, followed by transition to the long-term state (with probability a); or an incorrect response and no transition to the long-term state (with probability (1-g)(1-a)). In the latter case, there follows a make-up sequence of presentations, which are useless if Greeno's model holds because the item is trapped in the short-term state. Then intervening items are presented, and finally the item is tried once again. By this time the item is back in U (according to the approximation assumption), the process starts over and is repeated until one of the first two situations obtains. Thus

$$P(U_{k+1}|U_k) = g(1-a) \sum_{\nu=0}^{\infty} [(1-g)(1-a)]^{\nu} = \frac{g(1-a)}{a+g(1-a)} ,$$

as required.

Approximations Under the GFT

Theorem 3.3: A formula for $q_{k,n}$. In terms of the parameters of the GFT presented in Chapter 2, the conditional probability of an error on cycle n, given that the last error was on cycle k, is approximately given by the following formula.

$$(7)\quad q_{k,n} \doteq \begin{cases} \dfrac{(1-g)(1-a)[g(1-a)]^{n-2}}{[g(1-a)]^{n-2} + \dfrac{a}{1-g(1-a)}\,(1-[g(1-a)]^{n-2})} & \text{, for } k = 0,\ n \geq 2 \\[2ex] \dfrac{L'_k[g(1-a)]^{n-k-1}(1-g)}{1-L'_k\dfrac{1-g}{1-g(1-a)} + L'_k\dfrac{1-g}{1-g(1-a)}\,[g(1-a)]^{n-k-1}} & \text{, for } k \geq 1,\ n > k \end{cases}$$

where L_k = P (In state L|make-up sequence of length k just completed) and $L_k' = 1-L_k$.

Proof. The expression for $q_{0,n}$ will be developed first.

$$q_{0,n} = \frac{P(\text{no errors through cycle } n-1, \text{ error on cycle } n)}{P(\text{no errors through cycle } n-1)}$$

$$= \frac{(1-g)[g(1-a)]^{n-1}}{g[g(1-a)]^{n-2} + \sum_{\nu=1}^{n-2} ag[g(1-a)]^{\nu-1}}$$

$$= \frac{(1-g)(1-a)[g(1-a)]^{n-2}}{[g(1-a)]^{n-2} + \frac{a}{1-g(1-a)}\left(1-[g(1-a)]^{n-2}\right)}$$

as asserted.

In order to develop the expression for $q_{k,n}$ when $k \geq 1$, let the events A and B be defined as follows.

A = {error on cycle n}

B = {last error through cycle n-1 was on cycle k}.

Then

$$q_{k,n} = P(A|B) = \frac{P(A \cap B)}{P(B)} .$$

The event B will occur if there is an error on cycle k, unless the item is still not in state L at the end of the make-up sequence in cycle k and there is another error before cycle n. Hence

$$P(B) = \pi_{k,k}\left(1 - L_k'(1-g) \sum_{\nu=0}^{n-k-2} [g(1-a)]^{\nu}\right)$$

$$= \pi_{k,k}\left(1 - \frac{1-g}{1-g(1-a)} L_k' + \frac{1-g}{1-g(1-a)} [g(1-a)]^{n-k-1} L_k'\right) .$$

The event $A \cap B$ will occur if there is an error on cycle k, and a series of correct guesses and failures to go to state L on subsequent cycles, ended by an incorrect guess on cycle n. The probability of this event is given by

$$P(A \cap B) = \pi_{k,k} L'_k [g(1-a)]^{n-k-1}(1-g) .$$

Dividing this expression by the expression just derived for P(B) yields the formula for $q_{k,n}$ given in Equation 7.

As one might expect, the quantities remaining to be calculated are closely interrelated. It is necessary to know L'_k in order to use Equation 7 to compute $q_{k,n}$. It will soon be seen that it is necessary to know $e_{k,k+1}$ in order to find L'_k. Calculating $e_{k,k+1}$ involves the determination of the distribution of the waiting time for the next error. It will help keep repetition to a minimum if we can refer to some basic quantities involved in the several separate calculations. The breakdown of W_k given in Figure 3.1 suggests what these quantities might be. Define five new random variables as follows.

K_U = the number of presentations in criterion run on cycle k in state U.

K_S, K_L = the corresponding random variables for states S and L.

$E^*_{k,U}$ = the number of presentations in state U following an error, given that there will be another error on cycle k.

$E^*_{k,S}$ = the corresponding random variable for state S.

It turns out that the remaining calculations in this section will be expedited by considering the joint distributions of (K_U, K_S, K_L) and

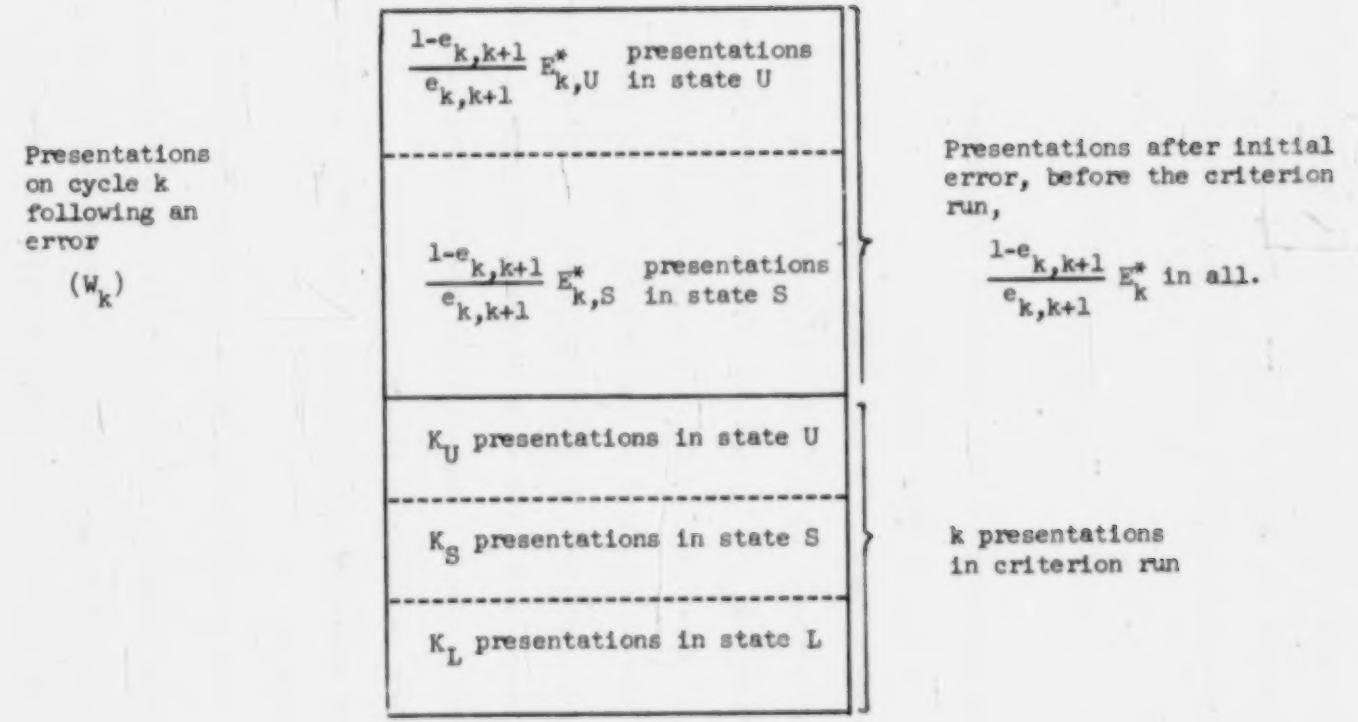
Presentations on cycle k following an error (W_k)
$\frac{1-e_{k,k+1}}{e_{k,k+1}} E^*_{k,U}$ presentations in state U
$\frac{1-e_{k,k+1}}{e_{k,k+1}} E^*_{k,S}$ presentations in state S
Presentations after initial error, before the criterion run, $\frac{1-e_{k,k+1}}{e_{k,k+1}} E^*_k$ in all.
K_U presentations in state U
K_S presentations in state S
K_L presentations in state L
k presentations in criterion run

Figure 3.1. Breakdown of presentations on cycle k after initial error.

$(E^*_{k,U}, E^*_{k,S})$, respectively. Note that

$$K_U + K_S + K_L = k \quad \text{and} \quad E^*_{k,U} + E^*_{k,S} = E^*_k .$$

The joint distribution of $(E^*_{k,U}, E^*_{k,S})$ is a conditional distribution given that another error is going to occur. For this reason the probability of the simple sequence of events that results in the event $\{E^*_{k,U} = \ell, E^*_{k,S} = m\}$ for some ℓ and m must be divided by $1-e_{k,k+1}$ in order to find $P(E^*_{k,U} = \ell, E^*_{k,S} = m)$. Figure 3.2 shows a classification of points (ℓ,m) having positive probability into three types such that the probability expressions are similar for points within a type. The figure applies to the special case k = 4. The reader can easily verify the following assertion if he bears in mind that presentations in a make-up sequence on cycle k are contiguous until there are k-1 presentations receiving correct responses, in which case a number of items intervene before the next presentation. Let $\gamma = g(1-a-b)$. Then

$$(1-e_{k,k+1})P(E^*_{k,U} = \ell,\ E^*_{k,S} = m)$$

$$(8) \qquad \doteq \begin{cases} \gamma^{\ell-1}(1-a-b)(1-g) & , \text{ for } \ell=1,\ldots,k-1;\ m=0; \\ \gamma^{k-1}(1-a)(1-g) & , \text{ for } \ell=k;\ m=0; \\ \gamma^{\ell-1}\, b(1-c)^{k-\ell}(1-g) & , \text{ for } \ell=1,\ldots,k-1;\ m=k-\ell . \end{cases}$$

Figure 3.3 gives a classification of points (ℓ,m,n) having similar expressions for $e_{k,k+1}\, P(K_U = \ell, K_S = m, K_L = n)$. A suitable approximation for this quantity is given by

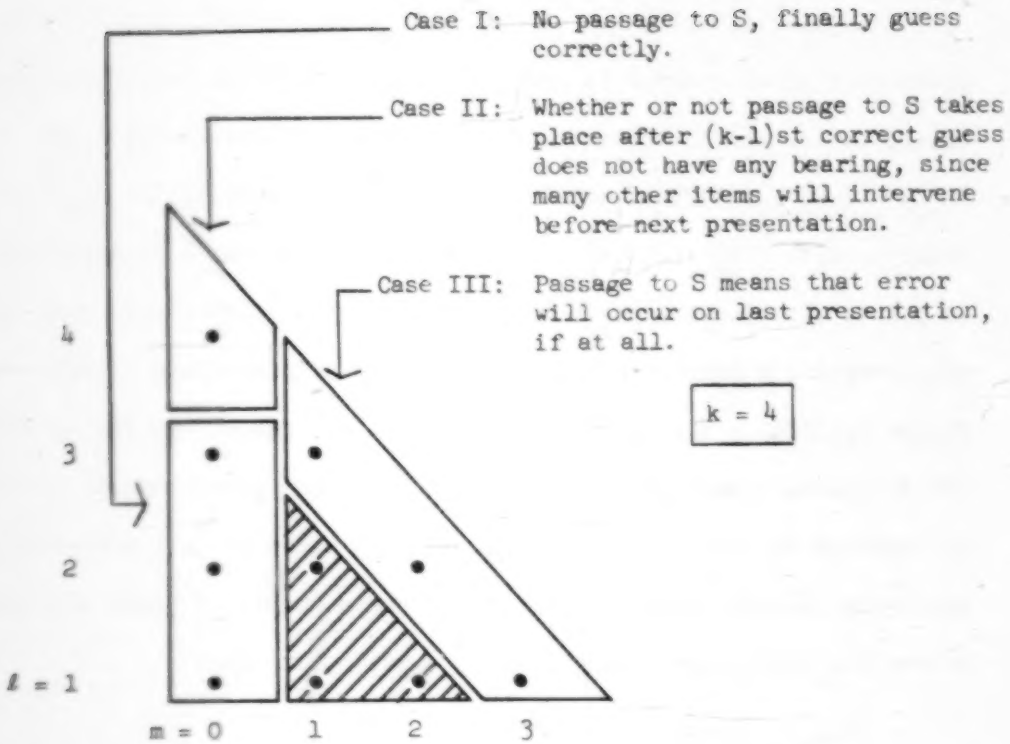
Case I: No passage to S, finally guess correctly.
Case II: Whether or not passage to S takes place after (k-1)st correct guess does not have any bearing, since many other items will intervene before next presentation.
Case III: Passage to S means that error will occur on last presentation, if at all.
k = 4
4
3
2
ℓ = 1
m = 0
1
2
3

Figure 3.2. Classification of points (ℓ,m) having similar expressions for $(1-e_{k,k+1})P(E^*_{k,U} = \ell, E^*_{k,S} = m)$, where $E^*_{k,U}$, $E^*_{k,S}$ are the numbers of presentations in the respective states in a run culminating in an error on cycle 4.

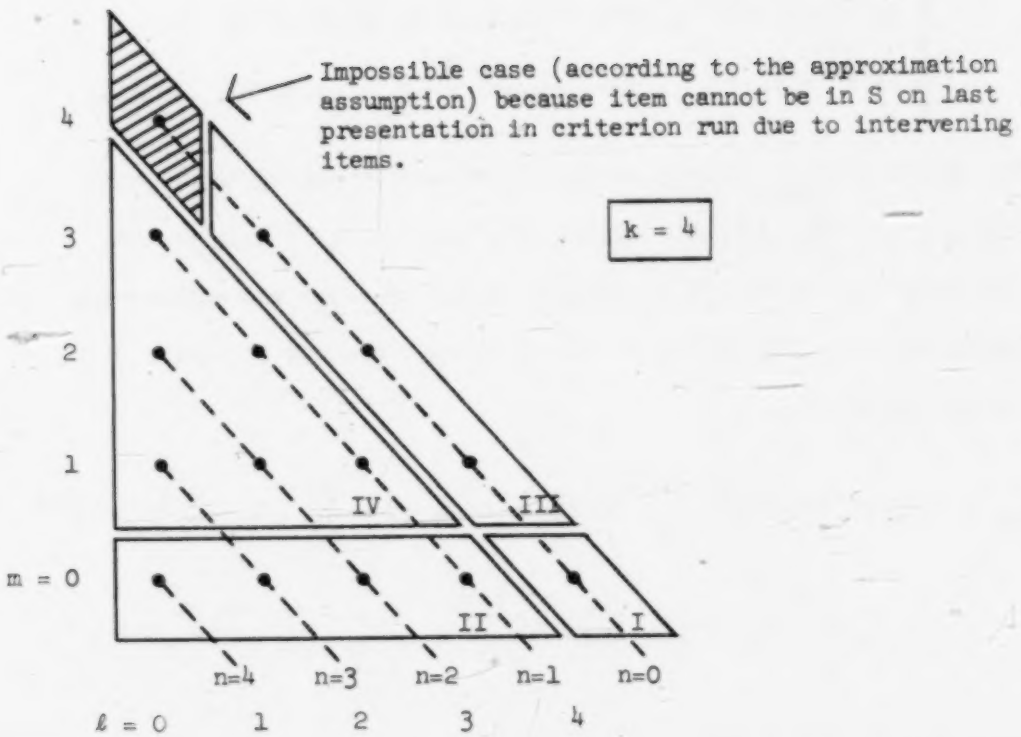

Figure 3.3. Classification of points (ℓ,m,n) having similar expressions for $e_{k,k+1}\, P(K_U = \ell, K_S = m, K_L = n)$. where K_U, K_S, K_L are the numbers of presentations in the respective states during the criterion run on cycle 4.

$$e_{k,k+1}\, P(K_U = \ell, K_S = m, K_L = n)$$

$$(9) \qquad \doteq \begin{cases} \gamma^{\ell-1} g(1-a) & , \text{ for } \ell=k;\ m=n=0; \\ a\gamma^{\ell} & , \text{ for } \ell=0,\ldots,k-1;\ m=0;\ n=k-\ell; \\ g\gamma^{\ell-1} b(1-c)^{k-\ell} & , \text{ for } \ell=1,\ldots,k-1;\ m=k-\ell;\ n=0; \\ \gamma^{k-m-n} b(1-c)^{m} c & , \text{ for } \ell=k-m-n;\ m=1,\ldots,k-n;\ n=1,\ldots,k-1. \end{cases}$$

The value of $e_{k,k+1}$ can be calculated either by adding the results of Equation 9 for all allowable values of (ℓ,m,n) or by adding the results for Equation 8 over all a possible values of (ℓ,m) and subtracting the outcome from 1. The latter course is simpler because it avoids a messy double summation.

$$(10) \qquad \begin{aligned} e_{k,k+1} &\doteq 1-(1-g)(1-a)\gamma^{k-1} - \sum_{\ell=1}^{k-1} \gamma^{\ell-1}(1-a-b)(1-g) - \sum_{\ell=1}^{k-1} (1-g)\gamma^{\ell-1} b(1-c)^{k-\ell} \\ &= 1-(1-g)\left[(1-a)\gamma^{k-1} + (1-a-b)\,\frac{1-\gamma^{k-1}}{1-\gamma} + \frac{b}{1-\frac{\gamma}{1-c}}\left[(1-c)^{k-1}-\gamma^{k-1}\right]\right] . \end{aligned}$$

This expression can be simplified significantly in some important special cases.

$$(11) \qquad e_{k,k+1} \doteq \begin{cases} 1 - b(1-g)(1-c)^{k-1} & , \text{ when } a+b = 1,\ c > 0; \\ 1 - \frac{(1-g)(1-a)}{1-\gamma}\,(1-\gamma^{k}) & , \text{ when } a+b < 1,\ c = 0; \\ 1 - b(1-g) & , \text{ when } a+b = 1,\ c = 0 . \end{cases}$$

<u>A formula for</u> L'_k. It was noted above that the formula for $q_{k,n}$ given in Equation 7 presumes knowledge of L'_k. For $k = 1$, $L_1 = a$, and $L'_1 = 1 - a$ because items are just presented once on the first cycle,

regardless of the correctness of the responses. In order to derive an estimate of L_k' for $k \geq 2$, define events A, B, and C as follows.

A = {Error on cycle k}

B = {Error on cycle k, cycle k completed with no further errors}

C = {Error on cycle k, cycle k completed with no further errors, but item fails to make transition to state L}.

The fact that $C \subset B \subset A$ and $P(A) > 0$ implies that

$$L_k' = P(C|B) = \frac{P(C|A)}{P(B|A)} .$$

The quantity $P(B|A)$ is $e_{k,k+1}$, which was just derived. The quantity $P(C|A)$ can be obtained by multiplying Equation 9 by 1-a and adding over those possibilities for which $K_L = 0$. That is,

$$e_{k,k+1} P(K_U = \ell, K_S = m, K_L = 0 \text{ and no transition to L takes place after correct guess on last response})$$

$$= (1-a)\, e_{k,k+1} P(K_U = 1, K_S = k-\ell, K_L = 0)$$

$$= \begin{cases} \gamma^{\ell-1} g(1-a)^2 & \text{for } \ell = k \\ g\gamma^{\ell-1} b(1-c)^{k-\ell}(1-a) & \text{for } \ell = 1,\ldots,k-1; \end{cases}$$

and

$$P(C|A) = \sum_{\ell=1}^{k} (1-a)\, e_{k,k+1} P(K_U = \ell, K_S = k-\ell, K_L = 0)$$

$$= \gamma^{k-1} g(1-a)^2 + \sum_{\ell=1}^{k-1} g\gamma^{\ell-1} b(1-c)^{k-\ell}(1-a)$$

$$= \gamma^{k-1} g(1-a)^2 + bg(1-a)(1-c)^{k-1} \sum_{\ell=1}^{k-1} \left(\frac{\gamma}{1-c}\right)^{\ell-1}$$

$$= \gamma^{k-1} g(1-a)^2 + \frac{bg(1-a)}{1-\frac{\gamma}{1-c}} [(1-c)^{k-1} - \gamma^{k-1}]$$

$$= g(1-a)[(1-a - \frac{b}{1-\frac{\gamma}{1-c}})\gamma^{k-1} + \frac{b}{1-\frac{\gamma}{1-c}} (1-c)^{k-1}] .$$

This formula also simplifies in important special cases:

$$P(C|A) = \begin{cases} g(1-a)b(1-c)^{k-1} & , \text{ when } a+b = 1,\ c > 0; \\ (1-a)\gamma^k + \frac{bg(1-a)}{1-\gamma} & , \text{ when } a+b < 1,\ c = 0; \\ bg(1-a) & , \text{ when } a+b = 1,\ c = 0 . \end{cases}$$

Finally, the formula for L'_k for $k \geq 2$ is

$$L'_k = \frac{P(C|A)}{e_{k,k+1}}$$

(12)

$$= \frac{g(1-a)[(1-a- \frac{b}{1-\frac{\gamma}{1-c}})\gamma^{k-1} + \frac{b}{1-\frac{\gamma}{1-c}} (1-c)^{k-1}]}{1-(1-g)[(1-a)\gamma^{k-1} + (1-a-b)\frac{1-\gamma^{k-1}}{1-\gamma} + \frac{b}{1-\frac{\gamma}{1-c}} [(1-c)^{k-1} - \gamma^{k-1}]]} .$$

In the special cases mentioned above, Equation 12 becomes

(13)

$$L'_k = \begin{cases} \frac{g(1-a)b(1-c)^{k-1}}{1-b(1-g)(1-c)^{k-1}} & , \text{ when } a+b = 1,\ c > 0; \\ \frac{(1-a)[(1-\gamma)\gamma^k + bg]}{1-\gamma-(1-g)(1-a)(1-\gamma^k)} & , \text{ when } a+b < 1,\ c = 0; \\ \frac{g(1-a)b}{1-b(1-g)} & , \text{ when } a+b = 1,\ c = 0 . \end{cases}$$

It might be noted that the formula for $q_{1,n}$ given in Equation 7 reduces to the formula for $q_{0,n}$ when 1-a is substituted for L_1'. Equation 13 can be used with Theorems 3.2 and 3.3 to compute estimates of $\pi_{k,k}$. Theorem 3.1 tells us how to compute EW_k (and hence EP_k) given $\pi_{k,k}$, $e_{k,k+1}$, and $\mu(E_k^*)$, the mean waiting conditional waiting time for another error following an error on cycle k, given that there is going to be another error. Approximations for all but the last of these quantities have been given above.

The distribution and expected value of E_k^*. Since $E_k^* = E_{k,U}^* + E_{k,S}^*$, we can use Equation 8 to write

$$(14)\quad (1-e_{k,k+1})\,P(E_k^* = \nu)$$

$$\doteq \begin{cases} (1-e_{k,k+1})\,P(E_{k,U}^* = \nu, E_{k,S}^* = 0), & \text{for } \nu = 1,\ldots,k-1; \\ (1-e_{k,k+1}) \sum_{m=0}^{k-1} P(E_{k,U}^* = k-m, E_{k,S}^* = m), & \text{for } \nu = k; \end{cases}$$

$$\doteq \begin{cases} \gamma^{\nu-1}(1-a-b)(1-g), & \text{for } \nu = 1,\ldots,k-1; \\ \gamma^{k-1}(1-a)(1-g) + \sum_{m=1}^{k-1} \gamma^{k-1-m} b(1-c)^m (1-g), & \text{for } \nu = k; \end{cases}$$

$$\doteq \begin{cases} \gamma^{\nu-1}(1-a-b)(1-g), & \text{for } \nu = 1,\ldots,k-1; \\ \gamma^{k-1}\left(1-a-\frac{b}{1-\frac{\gamma}{1-c}}\right)(1-g) + \frac{b(1-g)}{1-\frac{\gamma}{1-c}}(1-c)^{k-1}, & \text{for } \nu = k. \end{cases}$$

In the special cases, Equation 14 becomes

(15) $(1-e_{k,k+1})\ P(E_k^* = \nu)$

$$= \begin{cases} \begin{cases} b(1-g)(1-c)^{k-1} & , \text{ for } \nu = k; \\ 0 & , \text{ otherwise, when } a+b = 1,\ c > 0; \end{cases} \\ \begin{cases} \gamma^{\nu-1}(1-a-b)(1-g) & , \text{ for } \nu = 1,\ldots,k-1; \\ \gamma^{k-1}(1-a-\frac{b}{1-\gamma})(1-g) + \frac{b(1-g)}{(1-\gamma)} & , \text{ for } \nu = k \text{ when } a+b < 1,\ c = 0; \end{cases} \\ \begin{cases} b(1-g) & , \text{ for } \nu = k; \\ 0 & , \text{ otherwise, when } a+b = 1 \text{ and } c = 0. \end{cases} \end{cases}$$

Straight forward summation of series, omitted here for the sake of brevity, results in the following expression for $(1-e_{k,k+1})\mu(E_k^*)$.

(16) $$(1-e_{k,k+1})\mu(E_k^*)$$

$$= (1-a-b)(1-g)\left[\frac{1-k\gamma^{k-1}+(k-1)\gamma^k}{(1-\gamma)^2}\right]$$

$$+ k[\gamma^{k-1}(1-a-\frac{b}{1-\frac{\gamma}{1-c}})(1-g) + \frac{b(1-g)}{1-\frac{\gamma}{1-c}}(1-c)^{k-1}] .$$

The restriction that $c = 0$ does not in itself produce any significant simplification in Equation 16. The restriction that $a + b = 1$ does, however. In this case

(17) $$(1-e_{k,k+1})\mu(E_k^*) = \begin{cases} kt(1-g)(1-c)^{k-1} & , \text{ when } a+b = 1; \\ kb(1-g) & , \text{ when } a+b = 1 \text{ and } c = 0 . \end{cases}$$

Summary of the calculation of EP_k. If the expression for EW_k given in Equation 4 is substituted for EW_k in Equation 1, the result is

$$EP_k = 1 + \pi_{k,k} \left(k + \frac{(1-e_{k,k+1})\mu(E_k^*)}{e_{k,k+1}} \right) .$$

If the expressions derived above are substituted in the right-hand side of this equation, the result is a cumbersome expression for EP_k in terms of the model parameters. In general, this expression is unenlightening, so it will not be reproduced here. It should be noted, however, that it simplifies in the case a+b = 1 to the following:

$$EP_k \doteq 1 + \frac{k\pi_{k,k}}{1-b(1-g)(1-c)^{k-1}} .$$

Furthermore,

$$EP_k \doteq 1 + \frac{k\pi_{k,k}}{1-b(1-g)} , \quad \text{when} \quad a+b = 1 \quad \text{and} \quad c = 0 .$$

Other Operating Characteristics of the OEM Procedure

One of the chief reasons for studying the operating characteristics of the OEM procedure (under the assumption that some model in the GFT applies) is to find ways of modifying the procedure to get better instructional results. In a number of experiments it has been found that $a > c$. When this is the case in an instructional setting, one should maximize the cumulative number of presentations in state U. Thus, it is useful to consider how many of the P_k presentations of an item on cycle k are in states U, S, and L, respectively. This information may suggest modifications which would increase the proportion of presentations in state U.

Let $P_{k,U}$, $P_{k,S}$, and $P_{k,L}$ be the number of presentations in the respective states on cycle k. Let $\mu(K_U)$, $\mu(K_S)$, and $\mu(K_L)$ be the means

of K_U, K_S, and K_L, respectively, and let u_k, s_k, ℓ_k be the respective probabilities that an item is in U, S, or L on its first presentation on cycle k. Then

$$P_{k,U} + P_{k,S} + P_{k,L} = P_k ,$$

$$EP_{k,U} = u_k + \pi_{k,k} \left[\mu(K_U) + \frac{1-e_{k,k+1}}{e_{k,k+1}} \mu(E^*_{k,U})\right] ,$$

$$EP_{k,S} = s_k + \pi_{k,k} \left[\mu(K_S) + \frac{1-e_{k,k+1}}{e_{k,k+1}} \mu(E^*_{k,S})\right] ,$$

and

$$EP_{k,L} = \ell_k + \pi_{k,k} \mu(K_L) .$$

The task now is to find u_k, s_k, ℓ_k, $\mu(K_U)$, $\mu(K_S)$, $\mu(K_L)$, $(1-e_{k,k+1})\mu(E^*_{k,U})$, and $(1-e_{k,k+1})\mu(E^*_{k,S})$.

Finding $(1-e_{k,k+1})\mu(E^*_{k,U})$ and $(1-e_{k,k+1})\mu(E^*_{k,S})$. The procedure is very direct: first find $(1-e_{k,k+1})\mu(E^*_{k,S})$ by adding weighted terms using Equation 8; then find $(1-e_{k,k+1})\mu(E^*_{k,U})$ by subtracting the result from $(1-e_{k,k+1})\mu(E^*_k)$, which is given by Equation 16.

$$(18) \qquad (1-e_{k,k+1})\mu(E^*_{k,S}) = \sum_{m=1}^{k-1} m(1-e_{k,k+1})P(E^*_{k,S} = m)$$

$$\doteq \sum_{m=1}^{k-1} m\gamma^{k-m-1}b(1-c)^m(1-g)$$

$$\doteq b(1-g)\left(\frac{1-c}{\gamma}\right)\gamma^{k-1} \sum_{m=1}^{k-1} m\left(\frac{1-c}{\gamma}\right)^{m-1}$$

$$= b(1-g)\left(\frac{1-c}{\gamma}\right)\gamma^{k-1}\left[\frac{1-k\left(\frac{1-c}{\gamma}\right)^{k-1}+(k-1)\left(\frac{1-c}{\gamma}\right)^{k}}{\left(1-\frac{1-c}{\gamma}\right)^2}\right]$$

$$= \frac{b(1-g)}{(1-\frac{\gamma}{1-c})^2}\left\{\frac{\gamma}{1-c}\left[\gamma^{k-1}-(1-c)^{k-1}\right] + (k-1)(1-c)^{k-1}(1-\frac{\gamma}{1-c})\right\}.$$

Then subtracting Equation 18 from Equation 16, we have

$$(19)\quad (1-e_{k,k+1})\mu(E^*_{k,U}) = (1-e_{k,k+1})\mu(E^*_k) - (1-e_{k,k+1})\mu(E^*_{k,S})$$

$$\doteq \frac{(1-a-b)(1-g)}{(1-\gamma)^2}\left[1-k\gamma^{k-1}+(k-1)\gamma^k\right] + k(1-g)(1-a)\gamma^{k-1}$$

$$+ \frac{b(1-g)}{(1-\frac{\gamma}{1-c})^2}\left[(1-c)^{k-1}-\gamma^k\right] - \frac{(k-1)b(1-g)}{1-\frac{\gamma}{1-c}}\gamma^{k-1}.$$

As was the case with $(1-e_{k,k+1})\mu(E^*_k)$, these formulas can be simplified significantly only if $a+b = 1$. The results for that case are

$$(20)\quad (1-e_{k,k+1})\mu(E^*_{k,U}) \doteq \begin{cases} b(1-g)(1-c)^{k-1} & \text{, when } a+b = 1, \\ b(1-g) & \text{, when } a+b = 1 \text{ and } c = 0; \end{cases}$$

and

$$(1-e_{k,k+1})\mu(E^*_{k,S}) \doteq \begin{cases} (k-1)b(1-g)(1-c)^{k-1} & \text{, when } a+b = 1, \\ (k-1)b(1-g) & \text{, when } a+b = 1 \text{ and } c = 0. \end{cases}$$

Finding $\mu(K_U)$, $\mu(K_S)$, and $\mu(K_L)$. The approach to calculating these quantities is also direct. Expressions for $e_{k,k+1}\mu(K_S)$ and $e_{k,k+1}\mu(K_L)$ will be found first because the probability expressions for $K_S > 0$ and $K_L > 0$ given in Equation 9 involve only two of the four types pictured in Figure 3.3, while those for K_U involve all four types. Once $\mu(K_S)$ and $\mu(K_L)$ have been found, we know $\mu(K_U) = k-\mu(K_S)-\mu(K_L)$.

On the basis of Equation 9, we can write

$$(21)\qquad e_{k,k+1}\,P(K_L=n) \doteq a\gamma^{k-n} + \sum_{m=1}^{k-n} bc\gamma^{k-m-n}(1-c)^m$$

$$= \left(a - \frac{bc}{1-\frac{\gamma}{1-c}}\right)\gamma^{k-n} + \frac{bc}{1-\frac{\gamma}{1-c}}(1-c)^{k-n}\ ,\ \text{for } n=1,\ldots,k\ .$$

Therefore,

$$(22)\qquad e_{k,k+1}\mu(K_L) \doteq \left(a - \frac{bc}{1-\frac{\gamma}{1-c}}\right)\sum_{n=1}^{k} n\gamma^{k-n} + \frac{bc}{1-\frac{\gamma}{1-c}}\sum_{n=1}^{k} n(1-c)^{k-n}$$

$$= \begin{cases} ak\ , & \text{when } a+b=1,\ c=0; \\ a\left[\dfrac{k}{1-\gamma} - \dfrac{\gamma(1-\gamma^k)}{(1-\gamma)^2}\right], & \text{when } a+b<1,\ c=0; \\ ak + (1-c)bk - \dfrac{b(1-c)[1-(1-c)^k]}{c}, & \text{when } a+b=1,\ c>0; \\ \left(a - \dfrac{bc}{1-\frac{\gamma}{1-c}}\right)\left[\dfrac{k}{1-\gamma} - \dfrac{\gamma(1-\gamma^k)}{(1-\gamma)^2}\right] + \dfrac{bc}{1-\frac{\gamma}{1-c}}\left[\dfrac{k}{c} - \dfrac{(1-c)[1-(1-c)^k]}{c^2}\right], & \text{when } a+b<1,\ c>0\ . \end{cases}$$

The corresponding argument in the case of $e_{k,k+1}\mu(K_S)$ is that

$$(23)\qquad e_{k,k+1}\,P(K_S=m) = bg\gamma^{k-m-1}(1-c)^m + \sum_{n=1}^{k-m} bc\gamma^{k-m-n}(1-c)^m$$

$$= bg(1-c)^m\,\gamma^{k-m-1} + \frac{bc(1-c)^m}{1-\gamma}(1-\gamma^{k-m}),\ \text{for } m=1,\ldots,k-1;$$

therefore,

(24) $e_{k,k+1}\mu(K_S)$

$$\doteq \begin{cases} bg(k-1), & \text{when } a+b = 1,\ c = 0; \\ bg\left[\frac{k-1}{1-\gamma} - \frac{\gamma(1-\gamma^{k-1})}{(1-\gamma)^2}\right], & \text{when } a+b < 1,\ c = 0; \\ bg(1-c)^k(k-1) + \frac{1-c}{c}\, b[1-(1-c)^{k-1} - (k-1)(1-c)^{k-1}c], & \text{when } a+b = 1,\ c > 0; \\ b(\frac{g}{\gamma} - \frac{c}{1-\gamma})(\frac{1-c}{\gamma})[\gamma^k - k\gamma(1-c)^{k-1} + (k-1)(1-c)^k] & \\ \quad + \frac{b}{1-\gamma} \cdot \frac{1-c}{c}\,[1-(1-c)^{k-1} - (k-1)(1-c)^{k-1}c], & \text{when } a+b < 1,\ c > 0\,. \end{cases}$$

Approximations for $\mu(K_L)$, $\mu(K_S)$, and $\mu(K_U)$ can now be computed using Equations 22, 24 and the formulas for $e_{k,k+1}$ given in Equations 10 and 11. For example, in the simplest non-trivial case, where $a+b = 1$ and $c = 0$, they are given by

$$\mu(K_L) \doteq \frac{ka}{1-b(1-g)}\,,$$

$$\mu(K_S) \doteq \frac{(k-1)bg}{1-b(1-g)}\,,$$

and

$$\mu(K_U) \doteq \frac{bg}{1-b(1-g)}\,.$$

The values of u_k, e_k, and ℓ_k are needed in order to complete the state-by-state breakdown of the average number of presentations on cycle k. These are easily determined by the relationships $\pi_{k,k} = (1-g)u_k$,

$1-\pi_{k,k} \doteq \ell_k + gu_k$, and $s_k \doteq 0$. The fact that $s_k \doteq 0$ is a consequence of the approximation assumptions. These relationships imply that

$$u_k \doteq \frac{\pi_{k,k}}{1-g} \tag{25}$$

and

$$\ell_k \doteq 1 - \frac{\pi_{k,k}}{1-g}.$$

Formulas have been derived which provide for the calculation of the main quantities of interest when the OEM strategy is employed. We turn now to the calculation of analogous quantities when other strategies are used.

Operating Characteristics of Other Strategies

An Ideal Modification of the OEM Strategy

Most experiments to date concerned with the evaluation of the GPT framework suggest that parameter $a \geq c$. If the number of items being presented is large, say, greater than 20, then most of the items will not be in the short-term state at a given time. If it could be guaranteed that all items to be presented would not be in S, then the question of the optimal item to present would be reduced to the question of which item is most likely to still be in state U. The OEM would describe the learning process and the OEM strategy would be optimal. Such a modification of the OEM strategy would be ideal if it could be accomplished.

There are a number of ways one might approach this ideal. All items could be presented on a given cycle before any item receiving an error response was presented again. Those items receiving error responses

could then be presented in the standard cyclic fashion for the required number of make-up trials. Those items in the sublist receiving no errors in the remedial phase would be removed and the others would be presented once again and checked against the criterion, and so forth. The outcome of such a procedure should be that only a few items toward the end of the cycle would receive repeated presentations without a fairly large number of intervening items.

It is hard to say exactly how close an approach like the one just described would come to the ideal. The matter will not be pursued further here. But the operating characteristics of the ideal modification are easy to calculate, because they are just the characteristics derived in the last section, computed under the assumption that the OEM applies. The results will be stated here without proof, because they are based on well-known results for the OEM.

The mean number of presentations on cycle k. By Theorem 3.1 we know that

$$E(P_k) = 1 + \pi_{k,k} \left(k + \frac{1-e_{k,k+1}}{e_{k,k+1}} \mu(E_k^*)\right) .$$

Consider ℓ_j, the probability that an item is in state L at the end of the j^{th} cycle. Successive values of ℓ_j can be computed by the formula

$$\ell_{j+1} = \frac{\ell_j + (1-\ell_j)ag}{\ell_j + (1-\ell_j)g} \quad \text{where} \quad \ell_0 = 0 .$$

Then we have

$$\pi_{k,k} = (1-\ell_{k-1})(1-g) , \quad \text{for} \quad k \geq 1 .$$

By either letting b = 0 in Equation 16 or by straight forward argument from the properties of the OEM, it can be shown that

$$(26)\quad (1-e_{k,k+1})\mu(E_k^*) = \frac{(1-a)(1-g)}{[1-g(1-a)]^2}\left(1-[g(1-a)]^k\right) - k\,\frac{(1-a)(1-g)}{1-g(1-a)}\,[g(1-a)]^k .$$

Similarly, letting b = 0 in Equation 10 or proceeding by a direct argument yields

$$(27)\quad e_{k,k+1} = 1 - \frac{(1-g)(1-a)}{1-g(1-a)}\left(1-[g(1-a)]^{k-1}\right) .$$

It is interesting to compare these results with well-known results for the OEM. Let k become indefinitely large in Equations 26 and 27. Then we get

$$\lim_{k\to\infty} e_{k,k+1} = \frac{a}{1-g(1-a)} ,$$

which corresponds to the standard result for the probability of no more errors, following an error, when the OEM applies; and

$$\lim_{k\to\infty} (1-e_{k,k+1})\mu(E_k^*) = \frac{(1-a)(1-g)}{[1-g(1-a)]^2} .$$

If we let L be the trial of last error in an infinite sequence of presentations, then the mean of L is given by

$$EL = \frac{1-g}{a}\left(\frac{1}{1-g(1-a)}\right) .$$

Thus

$$\lim_{k\to\infty} \frac{(1-e_{k,k+1})\mu(E_k^*)}{e_{k,k+1}} = \frac{(1-a)(1-g)}{a[1-g(1-a)]} = (1-a)EL .$$

The quantity EL is the mean trial of last error, starting from state U. With probability a, an item moves to state L following an error, in which case there will certainly be no more errors. With probability 1-a, the process starts again from U.

Other operating characteristics of the ideal modification. By hypothesis, no items in the short-term state are ever presented under the modified strategy. Hence,

$$(1-e_{k,k+1})\mu(E^*_{k,U}) = (1-e_{k,k+1})\mu(E^*_k) ,$$

which has already been calculated. It is easy to show that

$$(28) \quad e_{k,k+1}\mu(K_U) = \frac{k[g(1-a)]^k[g(1-a)-[g(1-a)]^2-a]}{1-g(1-a)} + \frac{a[1-[g(1-a)]^k]}{[1-g(1-a)]^2} .$$

The breakdown of EP_k into the mean number of presentations in each state is as follows.

$$EP_{k,U} = 1-\ell_{k-1} + \pi_{k,k}\left[\mu(K_U) + \frac{1-e_{k,k+1}}{e_{k,k+1}}\mu(E^*_k)\right] ,$$

and

$$EP_{k,L} = \ell_{k-1} + \pi_{k,k}\mu(K_L) .$$

The asymptotic distribution of the cycle of last error. It may sometimes be of interest to consider what would happen if either the OEM procedure or its ideal modification were continued for a very large number of cycles. It is clear that sooner or later all items would be learned and there would be no further errors. In fact, the distribution of the cycle of last error can be expressed simply in terms of the $\pi_{k,k}$'s. Define $\pi_{k,\infty}$ and β_k as follows.

$$\pi_{k,\infty} = P(\text{last error is on cycle } k)$$

and

$$\beta_k = P(\text{no more errors} \mid \text{error on cycle } k) .$$

Then $\pi_{k,\infty} = \lim_{n\to\infty} \pi_{k,n}$, and $\pi_{k,\infty} = \pi_{k,k}\,\beta_k$. It follows from Equation 1c that

$$\beta_k = \lim_{n\to\infty} \prod_{\nu=1}^{n-k} (1-q_{k,k+\nu}) .$$

A more profitable way to look at β_k is to condition on whether an item is in state L or not following the make-up sequence on cycle k. Then we get

(29) $$\beta_k = L_k + L_k' \, P(\text{no more errors} \mid \text{in state U following cycle } k)$$

$$\doteq L_k + L_k' \sum_{\nu=k+1}^{\infty} [g(1-a)]^{\nu-k-1} ga$$

$$= L_k + L_k' \frac{ga}{1-g(1-a)}$$

$$= 1 - L_k' \frac{1-g}{1-g(1-a)} .$$

Therefore, the asymptotic probability that the last error is on cycle k is given approximately by

(30) $$\pi_{k,\infty} \doteq \pi_{k,:} \left(1 - L_k' \frac{1-g}{1-g(1-a)}\right) .$$

Operating Characteristics of the RC Procedure

For purposes of comparison, it is desirable to know the operating characteristics of the RC presentation procedure used in many experiments

on paired-associate learning. The usual approach used to derive theoretical predictions for this procedure when models like those in the GFT are being considered, is to multiply the learning matrix by an "average" forgetting matrix to get a single transition matrix summarizing the effects of learning following presentation and forgetting between presentations. Approximate theoretical predictions can then be made in terms of this single transition matrix. For the GFT, this matrix is given by

(31)
$$P = \begin{bmatrix} 1 & 0 & 0 \\ c & 1-c & 0 \\ a & b & 1-a-b \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1-f & f \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ c & (1-c)(1-f) & (1-c)f \\ a & b(1-f) & bf+(1-a-b) \end{bmatrix}$$

Calculation of the n-stage transition matrix, P^n, is facilitated by noting that the matrix P can be partitioned as follows, where B is a 2×2 matrix.

$$P = \begin{bmatrix} 1 & 0 \\ A & B \end{bmatrix}$$

It follows by simple algebra, that

(32)
$$P^n = \begin{bmatrix} 1 & 0 \\ \left(\sum_{i=1}^{n} B^{i-1}\right) A & B^n \end{bmatrix}$$

In the general case, there does not seem to be any particularly simple way to express B^n. Of course, it is very easy to carry out the multiplications numerically, so the lack of a simple expression causes no special difficulty. However, in the case $a+b = 1$, B^n does have a very simple form. If we let $\lambda = (1-f)(1-c) + bf$, it is easy to verify that

$$B^n = \lambda^{n-1} B .$$

This is so because λ is a characteristic root of B (and of P) and each row of B is a left characteristic vector corresponding to λ. Therefore, P^n can be written as

$$(33) \qquad P^n = \begin{bmatrix} 1 & 0 & 0 \\ 1-(1-c)\lambda^{n-1} & (1-c)(1-f)\lambda^{n-1} & (1-c)f\lambda^{n-1} \\ 1-(1-a)\lambda^{n-1} & (1-a)(1-f)\lambda^{n-1} & (1-a)f\lambda^{n-1} \end{bmatrix}.$$

In the general case, two positive values of λ are given by the formula

(34)

$$\lambda = \frac{(1-c)(1-f)+bf+(1-a-b) \pm \sqrt{[(1-c)(1-f)+bf+(1-a-b)]^2 - 4(1-c)(1-f)(1-a-b)}}{2} .$$

Equation 34 shows why the case $a+b = 1$ is special. A modification of Equation 33 using both λ's is useful in computing P^n in the general case, even though the theoretical formulas would be very messy in terms of the basic model parameters.

In the case $a+b = 1$, λ can be interpreted as the proportion of items currently in state S or state U which will still be in S or U following the next presentation.

If the starting state vector, [0, 0, 1], is postmultiplied by P^n, the result is the expected state vector after trial n. Thus

(35)
$$\ell_n = 1-(1-a)\lambda^{n-1}$$
$$s_n = (1-a)(1-f)\lambda^{n-1}, \text{ and}$$
$$u_n = (1-a)f\lambda^{n-1} .$$

Equation 35 can be used to derive all the operating characteristics one desires. The details will be discussed in the next chapter, in which the operating characteristics derived in this chapter will be used in numerical comparison of procedures.

CHAPTER IV

NUMERICAL COMPARISON OF PRESENTATION STRATEGIES

The purpose of this chapter is to compare the three presentation strategies we have been considering, using the formulas which were derived in the last chapter with specific parameter values to make predictions. Two kinds of questions are of particular interest. One concerns comparison of strategies given a certain set of parameter values. For example, how big is the difference between the RC procedure and the OEM procedure in terms of how many items are in the long-term retention state after a given number of presentations? How big is the difference between the OEM procedure and its ideal modification? Another kind of question of interest concerns how answers to the first kind of question vary as a function of parameter values. Do changes in the rate of transition from the unlearned to the long-term state affect the size of the differences between the OEM procedure and the RC procedure? In order to address these questions, operating characteristics of the three strategies have been computed for three different sets of parameter values. It might be helpful at this point to say a few words about the particular values that were chosen.

From the point of view of ease of calculation the best parameter values to choose satisfy the constraints of the model proposed by Greeno (1966): $c = 0$ and $a + b = 1$. That is, the probability of a presented item making a direct transition from the short-term to the long-term state is 0 and the probability is 1 that a presented item in the unconditioned state will make a transition either to the short-term or to the

long-term state. Expressions for operating characteristics of the OEM procedure are considerably simpler in this case than they are in general.

It was noted in the previous chapter that the OEM procedure leads to cyclic presentation of items, with repeated presentation of items receiving error responses on a given cycle. If Greeno's model holds, these massed presentations are useless. Immediately following the first presentation of an item on a given cycle the item is either in the long-term or short-term retention state, because $a + b = 1$. It is unnecessary to present it again if it is in the long-term state and it is useless to present it again immediately if it is in the short-term state, because $c = 0$. We want to examine the predictions of Greeno's model in some detail because we would expect them to differ sharply from the predictions of the OEM.

In contrast to Greeno's model the LS-2 model proposed by Atkinson and Crothers (1964) makes almost the same predictions for operating characteristics of the OEM procedure as the OEM itself. As in Greeno's model $a + b = 1$, but in the LS-2 model $c = a$. Since the probability of the presented item making a transition to the long-term state is the same whether the item is in the short-term or unconditioned state, the fact that $a + b = 1$ does not diminish the value of the massed presentations which occur under the OEM procedure. Greeno's model and the LS-2 model make the same predictions for the ideal modification of the OEM procedure and for the RC procedure because these predictions depend only on a, the transition probability between the unconditioned and the long-term state. For this reason it is unnecessary to calculate the operating characteristic for the LS-2 model directly unless one is interested in

the small, detailed differences between the LS-2 and OEM models under the OEM procedure.

Both cases mentioned so far satisfy the constraint $a+b = 1$. One of the interesting features of the data presented by Rumelhart (1967) was that the accuracy of predictions within the GFT framework could be significantly enhanced by allowing $a+b < 1$. This suggests that operating characteristics might be different for these two cases also. We will compare the prediction of two models which differ only with respect to whether $a+b = 1$ or $a+b < 1$. It will be seen that, contrary to our expectations, the differences in operating characteristics based on the two sets of parameters are minimal.

The reason the differences are small has to do with the most important variable influencing the relative performance of the strategies: the transition probability from the unlearned to the long-term state, a. When a is relatively large, as it is in Rumelhart's experiment, the differences between the three strategies are moderate and relatively insensitive to the values of the other parameters. When a is relatively small, the differences are pronounced and dependent on the values of the other parameters. These points will be expanded upon in an analysis of the detailed predictions for the three strategies using the three sets of parameter values given in Table 4.1.

Predictions of Greeno's Model When Learning is Slow

Atkinson and Crothers (1964) compared the fit of several models, including their LS-2 model, on eight different sets of experimental data. Of the eight experiments, the rate of learning was slowest in an experiment conducted by Hansen (1963) with four and five-year old nursery

Table 4.1

Three Sets of Parameter Values to be Used to Generate Theoretical Predictions of Operating Characteristics Under the Three Presentation Strategies

Case	Parameter values					Experiment
	a	b	c	f	g	
1	.129	.871	0	.844	.250	Four and five year old children. Hansen (1963)
2a	.410	.590	0	.954	.333	University undergraduates. Rumelhart (1967). Both cases fit data from same experiment. Case 2b relaxes constraint that a+b=1.
2b	.380	.360	0	.702	.333	

school children. Atkinson and Crothers report parameter estimates for the LS-2 model for this data. We are more interested in predictions for Greeno's model, for reasons described above. In order to adapt the parameter estimates for the LS-2 model to Greeno's model, we use the fact that if a' and f are parameters of the LS-2 model and $\underline{a}$ is the learning rate in Greeno's model, then Greeno's model with $\underline{a} = a'f$ and the same value of f will yield exactly the same predictions for the RC procedure as the LS-2 model. The parameter values for Case 1 given in 4.1 were obtained using this adjustment.

Before considering the operating characteristics predicted from these parameters, the reader may wish to review the breakdown of presentations on a given cycle under the OEM procedure which is summarized in Tables 4.2a and 4.2b. Table 4.2a gives general terms and their explicit expression in the case of Greeno's model. Table 4.2b identifies the terms used in the formulas in Table 4.2a.

It is worth noting that under the OEM procedure Greeno's model predicts that the average number of presentations in the unlearned state on a given cycle, $EP_{k,U}$, is a constant multiple of $\pi_{k,k}$, the probability of error on the cycle. The number of presentations in the short-term state, $EP_{k,S}$, on the other hand, depends on the product $(k-1)\pi_{k,k}$. The overall expected number of presentations on cycle k is 1 plus a constant multiple of $\pi_{k,k}k$. These observations provide a basis for determining whether or not the OEM procedure will be far from optimal under Greeno's model. Massed presentations are a real problem only on later cycles where the criterion run is long. If $\pi_{k,k}k$ converges to 0 relatively fast with increasing k, they are not a problem. If not, the OEM procedure will

Table 4.2a

General Formulas Giving a Breakdown of the Expected Number of Presentations of an Item on a Typical Cycle k, with Simplified Expressions Which Apply When Greeno's Model Holds*

	State			
Part of cycle	U	S	L	All states
First presentation	u_{k-1}	0	ℓ_{k-1}	1
Between first presentation and criterion run	$\pi_{k,k}\ \frac{1-e_{k,k+1}}{e_{k,k+1}}\ \mu(E^{e}_{k,U})$ $\doteq \frac{b(1-g)\pi_{k,k}}{1-b(1-g)}$	$\pi_{k,k}\ \frac{1-e_{k,k+1}}{e_{k,k+1}}\ \mu(E^{e}_{k,S})$ $\doteq \frac{(k-1)b(1-g)\pi_{k,k}}{1-b(1-g)}$	0	$\pi_{k,k}\ \frac{1-e_{k,k+1}}{e_{k,k+1}}\ \mu(E^{e}_{k})$ $= \frac{kb(1-g)\pi_{k,k}}{1-b(1-g)}$
Criterion run	$\pi_{k,k}\mu(K_U)$ $\doteq \frac{bg\pi_{k,k}}{1-b(1-g)}$	$\pi_{k,k}\mu(K_S)$ $\doteq \frac{(k-1)bg\pi_{k,k}}{1-b(1-g)}$	$\pi_{k,k}(K_L)$ $\doteq \frac{ka\pi_{k,k}}{1-b(1-g)}$	$\pi_{k,k}k$
Total	$EP_{k,U} \doteq u_{k-1} + \frac{b\pi_{k,k}}{1-b(1-g)}$	$EP_{k,S} \doteq \frac{(k-1)b\pi_{k,k}}{1-b(1-g)}$	$EP_{k,L} \doteq \ell_{k-1} + \frac{ka\pi_{k,k}}{1-b(1-g)}$	$EP_{k} \doteq \frac{k\pi_{k,k}}{1-b(1-g)}$

*Note: See key in Table 4.2b for identification of terms.

Table 4.2b

Key to Terms Used in Table 4.2a

---

u_{k-1}, ℓ_{k-1} = probability item is in state U (or L) following cycle k-1.

$\pi_{k,k}$ = probability of an error on the first response in cycle k.

$e_{k,k+1}$ = probability of no further errors on cycle k, following an error.

$\frac{1}{e_{k,k+1}}$ = expected number of errors on cycle k, given that at least one error occurs on the cycle.

$(1-e_{k,k+1})\mu(E^*_k)$ = expected waiting time for the next error on cycle k, in terms of number of responses following an error, given that another error is going to occur.

$(1-e_{k,k+1})\mu(E^*_{k,U}), (1-e_{k,k+1})\mu(E^*_{k,S})$ = breakdown of $(1-e_{k,k+1})\mu(E^*_k)$ by state.

$\mu(K_U), \mu(K_S), \mu(K_L)$ = expected number of presentations in the respective states on the criterion run on cycle k.

$EP_k, EP_{k,U}, EP_{k,S}, EP_{k,L}$ = expected number of presentations on cycle k, with breakdown by states.

---

be far from optimal. This can be seen in the present case, where learning is slow, by comparing the performance of the OEM procedure and its ideal modification.

The ideal modification of the OEM strategy serves two purposes in the comparisons to be made now. It presents an upper bound on how much the OEM strategy can be improved by simply manipulating the number of intervening items between presentations of an item on a cycle. It also gives indication of the discrepancy between the predictions of the OEM and of Greeno's model for the OEM strategy, since its operating characteristics are what the OEM would predict, with or without the modification. When learning is slow, this discrepancy is pronounced, as may be seen in Table 4.3.

The effect of the two procedures on $\pi_{K,K}$ is the same for the first two cycles, so $\pi_{3,3} = .410$ for both of them. But the OEM procedure requires 5.77 presentations per item to reduce error probability to this point, whereas the modified OEM procedure requires only 4.07 presentations. The size of the discrepancy increases for the next several cycles. Two more cycles under the modified procedure reduces $\pi_{K,K}$ to .04, a point that requires five additional cycles to reach under the OEM strategy. In terms of number of presentations per item, the comparison is 22.67 versus 10.90 presentations, a difference of more than 100%.

It is also interesting to compare the probability that an item is in the long-term state after a given number of cycles under the modified and unmodified OEM strategies with the corresponding probability for an item receiving the same number of presentations under the RC procedure. The difference between the modified OEM procedure and the RC procedure

Table 4.3

Predictions of Greeno's Model for Selected Operating Characteristics of the OEM, Modified OEM, and RC Procedures When Learning is Relatively Slow*

(a) OEM procedure

Cycle number k	Probability of error on cycle k	Expected cumulative number of presentations through cycle k	Probability item is in long-term state after k cycles under OEM procedure	Probability item is in long-term state after same number of presentations under RC procedure
1	.750	1.00	.129	.129
2	.653	5.77	.453	.497
3	.410	10.32	.656	.702
4	.258	14.29	.784	.812
5	.162	17.62	.864	.872
6	.102	20.38	.915	.907
7	.064	22.67	.947	.928
8	.040	24.60	.966	.943

(b) Modified OEM procedure

1	.750	1.00	.129	.129
2	.653	4.07	.453	.390
3	.410	8.37	.798	.627
4	.151	10.90	.948	.722
5	.039	12.34	.988	.764
6	.009	13.45	.997	.793
7	.003	14.48	.999	.816
8	.000	15.49	1.000	.836

*Note: Parameter values: a = .129, g = .25.

is dramatic in this respect, while the difference between the unmodified OEM and RC procedures is small and actually in the wrong direction for the first five cycles. See the last two columns of Table 4.3 for these comparisons.

Predictions of Greeno's Model When Learning is Rapid

Parameter values used to obtain predictions of Greeno's model when learning is rapid are given in Table 4.1, Case 2a. They are reported by Rumelhart (1967) to be the minimum chi-square estimates, computed by a grid search, for data from an experiment involving Stanford undergraduates.

One property of Greeno's model which has already been noted is that the probability of error on a cycle is the same for both the modified and unmodified procedures for the first three cycles. After the first three cycles, it drops more rapidly for the modified procedure. When learning is slow, this results in notable differences between the two procedures in terms of $\pi_{k,k}$, for $k > 3$. In the present case, learning is so rapid that there is little room for the $\pi_{k,k}$'s for $k > 3$ to differ, because they are all close to 0. Even though the $\pi_{k,k}$'s are close for the two procedures, it is conceivable that the procedures differ in terms of the number of presentations required to complete cycles. They do in this case, but only slightly. For example, it takes 3.30 presentations on the average to finish two cycles under the OEM procedure and 3.14 presentations under the modified procedure.

By referring to Table 4.4, the reader can see that the small differences between the modified and unmodified OEM procedures are typical of the differences that can be considered. All the differences favor the modified OEM procedure, as they must, but none of the differences

Table 4.4

Predictions of Greeno's Model for Selected Operating Characteristics of the OEM, Modified OEM, and RC Procedures When Learning is Relatively Rapid*

(a) OEM procedure

Cycle number k	Probability of error on cycle k	Expected cumulative number of presentations through cycle k	Probability item is in long-term state after k cycles under OEM procedure	Probability item is in long-term state after same number of presentations under RC procedure
1	.667	1.00	.410	.410
2	.393	3.30	.809	.807
3	.128	4.93	.938	.915
4	.041	6.20	.980	.955
5	.013	7.31	.992	.974
6	.005	8.36	.998	.984

(b) Modified OEM procedure

1	.667	1.00	.410	.410
2	.393	3.14	.809	.793
3	.128	4.67	.957	.902
4	.029	5.82	.991	.945
5	.006	6.85	.998	.967
6	.001	7.86	1.000	.980

*Note: Parameter values: $a = .41$, $g = .33$.

are large. Examination of the last two columns of Table 4.4 also reveals that when learning is rapid, Greeno's model predicts that the differences between the OEM procedure, modified or unmodified, and the RC procedure will be slight.

Predictions of a More General Model When Learning is Rapid

The parameters of Case 2b in Table 4.1 are what Rumelhart obtained for the data just described when he relaxed the requirement that a+b = 1. Predictions of operating characteristics of the OEM procedure made by Greeno's model and the more general model are compared in Table 4.5. The differences in the predictions are very slight indeed. In general, one would expect there to be a difference in the predictions the two models make regarding the expected number of presentations in the short-term state per cycle. In the present case the model for which $a+b < 1$ predicts about a third fewer presentations in the short-term state than does Greeno's model, but the rate of learning is so great that the number of these presentations is predicted to be small by both models.

Summary of the Relative Performance of the Three Strategies

It might be helpful to review some properties of the special cases of the GPT we have been considering before summarizing the results. No cases have been examined for which $c > a$. There are two reasons for this omission: first, no experiments have been reported for which $c > a$, at least to the author's knowledge; second, the case $c > a$ radically modifies what is desirable in a strategy, because it is then desirable to present items which are in the short-term state. The question of good presentation strategies for this case would be interesting in itself if situations arise where it applies. Among cases where $c \leq a$, we have

Table 4.5

Comparison of Predictions of Greeno's Model and a Model that Permits $a + b < 1$

Cycle number k	Probability that item is in state L at start of cycle k		Expected number of cumulative presentations through cycle k		Expected number of presentations of items in state S on cycle k	
	a+b=1	a+b<1	a+b=1	a+b<1	a+b=1	a+b<1
1	---	---	1.00	1.00	.00	.00
2	.41	.38	3.30	3.37	.38	.27
3	.81	.81	4.93	4.99	.25	.18
4	.94	.94	6.20	6.24	.12	.08
5	.98	.98	7.31	7.33	.05	.03
6	.99	.99	8.36	8.37	.02	.01

*Note: For Greeno's model parameters are a = .41, b = .59.
For the more general model a = .38, b = .36.

considered or can guess what would be predicted for models having extreme values of the three parameters. That is, we have some basis for saying what will happen for a large and small, for b = 0 and b = 1-a, for c = 0 and c = a. The predictions are as follows.

1. When a is large, differences on other parameters are not important. The operating characteristics of the three strategies are such that the modified OEM strategy has a slight advantage over the OEM strategy and the OEM has a slight advantage over the RC strategy. If a = c, the modified and unmodified OEM are practically identical.
2. When a is small and c = 0, the modified OEM procedure is far better than the other two. In this circumstance, the RC procedure may even be slightly superior to the OEM procedure.
3. When a is small and c = a, the modified OEM is much better than the RC procedure, but not much better than the unmodified OEM procedure.
4. Regarding the importance of parameter b, we may say:
   a. It is not important when a is large.
   b. If a is small and c = 0, b determines the relative performance of the OEM and modified OEM procedures. The worst case for the OEM procedure is when b = 1-a and the best case when b = 0. (In the latter case, the OEM and modified OEM operating characteristics are identical.)
   c. If a is small and c = a, the size of b is of little importance.

Some of the qualitative conclusions suggested here could be deduced heuristically without calculating the operating characteristics in numerical terms. The numerical calculations serve to transform the vague generalizations which could be made without them into assertions whose meaning can be made as precise and detailed as one wants.

CHAPTER V

CONCLUDING DISCUSSION

Attempts to deduce instructional implications from psychological theories or empirical generalizations may be crudely classified as belonging to one of two types. One type of deduction is very informal, perhaps, but not necessarily because the relationship on which it is based is only loosely formulated. Practically all deductions of implications for instructional practice were of this type until ten years ago. At that time, the success of some very explicit mathematical theories of simple learning processes led a few investigators to try more formal derivation of instructional strategies. Because the explicit mathematical statement of the consequences of instructional acts makes it possible to formulate the question of optimal instruction policy in completely unambiguous terms, it is natural to seek the answer to this question. The study undertaken in this paper is closer in spirit to this latter type of approach, but it does involve what some might regard as regressive elements of the first approach.

It was argued in the second chapter with regard to the question of item presentation strategy, that the globally optimal strategy corresponding to the GFT is too complicated to be of central interest. But what is of interest if the optimal strategy is not? Surely, if a reasonable model of the process of learning items exists, it should be possible to use it to make judgments about presentation strategies, even if it is not practical to work with the globally optimal strategy based on that model. The problem is that the bases for the judgments may come to

depend somewhat on the biases and preferences of the individual investigator. For example, investigator A may argue for the strategy maximizing immediate gain while investigator B pushes a modification of the OEM strategy, both justifying their choice on the basis of the GFT. Some theorists would regard this as an unpleasantly awkward situation; others would see nothing wrong with it. Tukey (1962), for example, has stated that the question of which statistical procedure is optimal in a given situation does not interest him until he knows of four sensible alternatives that have demonstrably different properties. At that point, lack of a criterion for choosing between the alternatives becomes a concern. It may happen that none of the alternatives is globally optimal, but one or more of them is very nearly optimal. If a theoretical analysis could identify such a situation when it occurs it would be very helpful, even if the analysis does not yield an optimal procedure.

The descriptive analysis of three presentation procedures under GFT assumptions given in Chapters III and IV provides some basis for saying when a strategy is nearly optimal. When learning is very rapid, for example, both the RC and the OEM strategies are very nearly optimal, independent of the exact parameter values. When learning is extremely slow, the RC procedure is poor, and how bad the OEM procedure is depends very much on the exact parameter values. These theoretical results are consistent with the results of the few empirical studies that have been done. Whether or not they will hold up under more direct experimental scrutiny is an open question.

There are a number of limitations imposed by the scope of this study which would need to be considered before applying the conclusions

in a particular list learning situation. It has been assumed that the GFT adequately describes the process governing learning and retention; that the items have neutral transfer value with respect to each other; that all the items are unknown at the start of instruction; that corresponding learning parameters are equal from one item to the next; and that the reward structure can be taken to be a simple function of the overall probability of correct response at the end of instruction. One or more of these assumptions are very likely to be violated in practice. Violations may or may not have damaging consequences for a given strategy. There are some relevant studies that relate to some of these consequences. Let us review some of them now.

The adequacy of the GFT framework. It is almost certain that the GFT framework could be shown to be an oversimplified account of the process of learning and retention. The phenomena of human information processing are now being studied with particular intensity. At times it seems as though important new developments are appearing monthly. In a climate of such intense experimental and theoretical inquiry all bets are off concerning the adequacy of any simple model. One aspect of the GFT which is suspect concerns its representation of what happens to an item which is not presented on a given trial. The GFT assumes that no learning takes place in this situation. But suppose a subject surreptitiously rehearses an item for a few trials after it has been presented. The GFT assumes that transitions to the long-term state could not take place via such a process. In fact, in a very successful model of human memory proposed by Atkinson and Shiffrin (1968), such a rehearsal process plays a central role. It is beyond the scope of this

paper to guess what the implications of other models of learning and memory might be with regard to item presentation strategy. But it is important to note that the GFT is not the only way that the memory processes can be modeled. For purposes of this study it suffices that it is a reasonable way to model the process.

Heterogeneity of items. The assumptions that items are all unknown at the start of instruction and that their learning parameters are equal from item to item are almost certain to be violated unless extraordinary measures are taken to insure that they hold. The necessity for such measures notably reduces the general applicability of the procedures requiring them. It may be that a procedure will be reasonably robust with respect to minor violations of homogeneity. Calfee (1970), for example, carried out some numerical calculations which suggest that this is the case for the OEM procedure, provided the other OEM assumptions are satisfied. We might argue, by analogy, that a suitably modified OEM procedure would stand up pretty well under minor deviations from item homogeneity, provided the other assumptions of the GFT hold. If item heterogeneity is extensive, however, a couple of studies have shown that parameter-dependent strategies which take these item differences into account will out-perform the OEM procedure. See Laubsch (1969), and Atkinson and Paulson (1972).

The importance of item heterogeneity has been noted; it should also be noted that in order to implement a parameter-dependent strategy the key problem is to find suitable parameter estimates. These estimates must separate subject and item differences, so we are led indirectly to a consideration of individual differences. The consideration of subject

differences required has an interesting twist to it: it is not enough to estimate the state of knowledge of a subject; one must also measure in a fairly "direct" sense the subject's ability to learn. These measurements may have some important implications for the concept of intelligence and its assessment. In a symposium on the nature of intelligence, Hunt (1972) described some exploratory studies of persons of above average intelligence who could be classified into two groups on the basis of being more quantitatively than verbally oriented, and vice versa. These subjects were given continuous memory tasks like the one described by Atkinson and Shiffrin (1968), and model parameters were estimated for their model. Consistent individual differences between subjects were found; these differences were meaningfully related to differences in their independently determined profiles of ability. Parameter-dependent strategies should be of continuing practical and theoretical interest.

Other crucial assumptions. It is patently clear for some kinds of curriculum material that some ways of sequencing the material make sense and others do not. When there is a natural sequence or hierarchy in the material to be learned, good presentation strategies must take them into account. Such strategies are beyond the scope of this study. Also beyond the scope of this study are situations where the items to be learned are differentially weighted in terms of their importance. Smallwood (1970) has considered this kind of problem, using an OEM theoretical framework which allows for item heterogeneity.

Final Remark

It should be apparent from the preceding discussion that the item selection problem is not a single problem with a single solution, but

rather is a family of problems representing a wide range of educational situations in which the question of optimal procedure is open. The study described in this paper has addressed one of these problems. Other studies examining some of the other problems which remain unresolved are in progress. It is the goal of all of these studies to develop general methods which may be used to attack more complex optimization problems. Three universal aspects of problems of optimizing instruction are emphasized: (1) the development of an adequate description of the learning process, (2) the assessment of costs and benefits associated with possible instructional actions and states of learning, and (3) the derivation of optimal strategies based on the goals set for the student.

The format of the list learning task is simple enough that all three aspects of optimization problems mentioned above can be subjected to detailed experimental and theoretical analysis. In addition to research described in this paper, instructional strategies which explicitly take individual differences into account are now being studied. There should be studies in the near future utilizing organizational features of the material to be learned in constructing optimal strategies. While the direct implications which can be drawn from such formal optimization studies of list learning are necessarily limited, the fact that many prototypical educational problems remain unresolved even within this restricted context justifies continued expenditure of effort at this level.

REFERENCES

Atkinson, R. C., & Crothers, E. J. A comparison of paired-associate learning models having different acquisition and retention axioms. Journal of Mathematical Psychology, 1964, 2, 285-315.

Atkinson, R. C., & Paulson, J. A. An approach to the psychology of instruction. Psychological Bulletin, 1972, 78, 49-61.

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence and J. T. Spence (Eds.), The psychology of learning and motivation: Advances in research and theory, Vol. II. New York: Academic Press, 1968.

Bernbach, H. A. A forgetting model for paired-associate learning. Journal of Mathematical Psychology, 1965, 2, 128-144.

Bower, G. H. Application of a model to paired-associate learning. Psychometrika, 1961, 26, 255-280.

Calfee, R. C. The role of mathematical models in optimizing instruction. Scientia: Revue Internationale de Synthèse Scientifique, 1970, 105, 1-25.

Dear, R. E., Silberman, H. F., Estavan, D. P., & Atkinson, R. C. An optimal strategy for the presentation of paired-associate items. Behavioral Science, 1967, 12, 1-13.

Feller, W. An introduction to probability theory and its applications. Vol. II. New York: John Wiley, 1966.

Greeno, J. G. Paired-associate learning with massed and distributed repetition of items. Journal of Experimental Psychology, 1964, 67, 286-295.

Greeno, J. G. Paired-associate learning with short-term retention: Mathematical analysis and data. Prepublication, 1966. Cited by Rumelhart (1967).

Hansen, D. N. Paired-associate learning with young children. Unpublished doctoral dissertation, Stanford University, 1963.

Hellyer, S. Supplementary report: Frequency of stimulus presentation and short-term decrement in recall. Journal of Experimental Psychology, 1962, 64, 650.

Hunt, E. Paper presented in a symposium on the theory of intelligence at the Western Psychological Association convention in Portland, Oregon, April 1972.

Karush, W., & Dear, R. E. Optimal stimulus presentation strategy for a stimulus sampling model of learning. Journal of Mathematical Psychology, 1966, 3, 19-47.

Laubsch, J. H. An adaptive teaching system for optimal item allocation. Technical Report 151, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1969.

Rumelhart, D. E. The effects of interpresentation intervals on performance in a continuous paired-associate task. Technical Report 116, Institute for Mathematical Studies in the Social Sciences, Stanford University, 1967.

Smallwood, R. D. The analysis of economic teaching strategies for a simple learning model. Journal of Mathematical Psychology, 1971, 8, 285-301.

Tukey, J. W. The future of data analysis. The Annals of Mathematical Statistics, 1962, 33, 1-67.

DISTRIBUTION LIST

Navy

4 Dr. Marshall J. Farr, Director
Personnel & Training Research Programs
Office of Naval Research
Arlington, VA 22217

1 Director
ONR Branch Office
495 Summer Street
Boston, MA 02210
ATTN: C. M. Harsh

1 Director
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101
ATTN: E. E. Gloye

1 Director
ONR Branch Office
536 South Clark Street
Chicago, IL 60605
ATTN: M. A. Bertin

1 Office of Naval Research
Area Office
207 West 24th Street
New York, NY 10011

6 Director
Naval Research Laboratory
Code 2627
Washington, DC 20390

12 Defense Documentation Center
Cameron Station, Building 5
5010 Duke Street
Alexandria, VA 22314

1 Chairman
Behavioral Science Department
Naval Command and Management Division
U.S. Naval Academy
Luce Hall
Annapolis, MD 21402

1 Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 38054
ATTN: Dr. G. D. Mayo

1 Chief of Naval Training
Naval Air Station
Pensacola, FL 32508
ATTN: Capt. Allen E. McMichael

1 LCDR Charles J. Theisen, Jr., MSC, USN
4024
Naval Air Development Center
Warminster, PA 18974

1 Commander
Naval Air Reserve
Naval Air Station
Glenview, IL 60026

1 Commander
Naval Air Systems Command
Department of the Navy
AIR-413C
Washington, DC 20360

1 Mr. Lee Miller (AIR-413E)
Naval Air Systems Command
5600 Columbia Pike
Falls Church, VA 22042

1 Dr. Harold Booher
NAVAIR 415C
Naval Air Systems Command
5600 Columbia Pike
Falls Church, VA 22042

1 Capt. John F. Riley, USN
Commanding Officer
U.S. Naval Amphibious School
Coronado, CA 92155

1 Special Assistant for Manpower
OASN (M&RA)
The Pentagon, Room 4E794
Washington, DC 20350

1 Dr. Richard J. Niehaus
Office of Civilian Manpower Management
Code 06A
Department of the Navy
Washington, DC 20390

1 CDR Richard L. Martin, USN
COMFAIRMIRAMAR F-14
NAS Mirsmar, CA 92145

1 Research Director, Code 06
Research and Evaluation Department
U.S. Naval Examining Center
Great Lakes, IL 60088
ATTN: C. S. Winiewicz

1 Chief
Bureau of Medicine and Surgery
Code 413
Washington, DC 20372

1 Program Coordinator
Bureau of Medicine and Surgery (Code 71G)
Department of the Navy
Washington, DC 20372

1 Commanding Officer
Naval Medical Neuropsychiatric
Research Unit
San Diego, CA 92152

1 Dr. John J. Collins
Chief of Naval Operations (OP-987F)
Department of the Navy
Washington, DC 20350

1 Technical Library (Pers-11B)
Bureau of Naval Personnel
Department of the Navy
Washington, DC 20360

1 Technical Director
Naval Personnel Research and
Development Center
San Diego, CA 92152

1 Commanding Officer
Naval Personnel Research and
Development Center
San Diego, CA 92152

1 Superintendent
Naval Postgraduate School
Monterey, CA 92940
ATTN: Library (Code 2124)

1 Mr. George N. Graine
Naval Ship Systems Command
(SHIPS 03H)
Department of the Navy
Washington, DC 20360

1 Technical Library
Naval Ship Systems Command
National Center, Building 3
Room 3S08
Washington, DC 20360

1 Commanding Officer
Service School Command
U.S. Naval Training Center
San Diego, CA 92133
ATTN: Code 303

1 Chief of Naval Training Support
Code N-21
Building 45
Naval Air Station
Pensacola, FL 32508

1 Dr. William L. Maloy
Principal Civilian Advisor
for Education and Training
Naval Training Command, Code 01A
Pensacola, FL 32508

1 Mr. Arnold Rubinstein
Naval Material Command (NMAT-03424)
Room 820, Crystal Plaza No. 6
Washington, DC 20360

Army

1 Commandant
U.S. Army Institute of Administration
Fort Benjamin Harrison, IN 46216
ATTN: EA

1 Armed Forces Staff College
Norfolk, VA 23511
ATTN: Library

1 Director of Research
U.S. Army Armor Human Research Unit
Building 2422, Morade Street
Fort Knox, KY 40121
ATTN: Library

1 U.S. Army Research Institute for the Behavioral and Social Sciences
1300 Wilson Boulevard
Arlington, VA 22209

1 Commanding Officer
USACDC - PASA
Ft. Benjamin Harrison, IN 46249
ATTN: LTC Montgomery

1 Dr. John L. Kobrick
Military Stress Laboratory
U.S. Army Research Institute of Environmental Medicine
Natick, MA 01760

1 Commandant
United States Army Infantry School
Fort Benning, GA 31905
ATTN: ATSIN-H

1 U.S. Army Research Institute
Commonwealth Building, Room 239
1300 Wilson Boulevard
Arlington, VA 22209
ATTN: Dr. R. Dusek

1 Mr. Edmund F. Fuchs
U.S. Army Research Institute
1300 Wilson Boulevard
Arlington, VA 22209

1 Chief, Unit Training and Educational Technology Systems
U.S. Army Research Institute for the Behavioral and Social Sciences
1300 Wilson Boulevard
Arlington, VA 22209

1 Commander
U.S. Theater Army Support Command, Europe
APO New York 09058
ATTN: Asst. DCSPER (Education)

1 Dr. Stanley L. Cohen
Work Unit Area Leader
Organizational Development Work Unit
Army Research Institute for Behavioral and Social Science
1300 Wilson Boulevard
Arlington, VA 22209

1 Dr. Leon H. Nawrocki
U.S. Army Research Institute
Rosslyn Commonwealth Building
1300 Wilson Boulevard
Arlington, VA 22209

Air Force

1 Headquarters, U.S. Air Force
Chief, Personnel Research and Analysis Division (AF/DPSY)
Washington, DC 20330

1 Research and Analysis Division
AF/DPXYR, Room 4C200
Washington, DC 20330

1 AFHRL/AS (Dr. G. A. Eckstrand)
Wright-Patterson AFB
Ohio 45433

1 AFHRL (AST/Dr. Ross L. Morgan)
Wright-Patterson AFB
Ohio 45433

1 AFHRL/MD
701 Prince Street
Room 200
Alexandria, VA 22314

1 AFOSR (NL)
1400 Wilson Boulevard
Arlington, VA 22209

1 Commandant
USAF School of Aerospace Medicine
Aeromedical Library (SUL-4)
Brooks AFB, TX 78235

1 CAPT Jack Thorpe, USAF
Department of Psychology
Bowling Green State University
Bowling Green, OH 43403

1 Headquarters, Electronic Systems Division
LG Hanscom Field
Bedford, MA 01730
ATTN: Dr. Sylvia R. Mayer/MCIT

Marine Corps

1 COL George Caridakis
Director, Office of Manpower Utilization
Headquarters, Marine Corps (AO1H)
MCB
Quantico, VA 22134

1 Dr. A. L. Slafkosky
Scientific Advisor (Code Ax)
Commandant of the Marine Corps
Washington, DC 20380

1 Mr. E. A. Dover
Manpower Measurement Unit (Code AO1M-2)
Arlington Annex, Room 2413
Arlington, VA 20370

Coast Guard

1 Mr. Joseph J. Cowan, Chief
Psychological Research Branch (P-1)
U.S. Coast Guard Headquarters
400 Seventh Street, SW
Washington, DC 20590

Other DOD

1 Lt. Col. Austin W. Kibler, Director
Human Resources Research Office
Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

1 Dr. Helga Yeich, Director
Program Management, Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

1 Mr. William J. Stormer
DOD Computer Institute
Washington Navy Yard
Building 175
Washington, DC 20374

1 Mr. Ralph R. Canter
Director for Manpower Research
Office of Secretary of Defense
The Pentagon, Room 3C980
Washington, DC 20301

Other Government

1 Office of Computer Information
Institute for Computer Sciences and Technology
National Bureau of Standards
Washington, DC 20234

Miscellaneous

1 Dr. Scarvia Anderson
Executive Director for Special Development
Educational Testing Service
Princeton, NJ 08540

1 Dr. Richard C. Atkinson
Department of Psychology
Stanford University
Stanford, CA 94305

1 Dr. Bernard M. Bass
Management Research Center
University of Rochester
Rochester, NY 14627

1 Mr. Edmund C. Berkeley
Berkeley Enterprises, Inc.
815 Washington Street
Newtonville, MA 02160

1 Dr. David G. Bowers
University of Michigan
Institute for Social Research
P.O. Box 1248
Ann Arbor, MI 48106

1 Mr. H. Dean Brown
Stanford Research Institute
333 Ravenswood Avenue
Menlo Park, CA 94025

1 Mr. Michael W. Brown
Operations Research, Inc.
1400 Spring Street
Silver Spring, MD 20910

1 Dr. Ronald P. Carver
American Institutes for Research
8555 Sixteenth Street
Silver Spring, MD 20910

1 Century Research Corporation
4113 Lee Highway
Arlington, VA 22207

1 Dr. Kenneth E. Clark
University of Rochester
College of Arts and Sciences
River Campus Station
Rochester, NY 14627

1 Dr. Allan M. Collins
Bolt Beranek and Newman
50 Moulton Street
Cambridge, MA 02138

1 Dr. René V. Dawis
University of Minnesota
Department of Psychology
Minneapolis, MN 55455

2 ERIC
Processing and Reference Facility
4833 Rugby Avenue
Bethesda, MD 20014

1 Dr. Victor Fields
Department of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Edwin A. Fleishman
American Institutes for Research
8555 Sixteenth Street
Silver Spring, MD 20910

1 Dr. Robert Glaser, Director
Learning Research and Development
Center
University of Pittsburgh
Pittsburgh, PA 15213

1 Dr. Albert S. Glickman
American Institutes for Research
8555 Sixteenth Street
Silver Spring, MD 20910

1 Dr. Duncan N. Hansen
Center for Computer-Assisted
Instruction
Florida State University
Tallahassee, FL 32306

1 Dr. Henry J. Hamburger
University of California
School of Social Sciences
Irvine, CA 92664

1 Dr. Richard S. Hatch
Decision Systems Associates, Inc.
11428 Rockville Pike
Rockville, MD 20852

1 Dr. M. D. Havron
Human Sciences Research, Inc.
Westgate Industrial Park
7710 Old Springhouse Road
McLean, VA 22101

1 Human Resources Research Organization
Division No. 3
P.O. Box 5787
Presidio of Monterey, CA 93940

1 Human Resources Research Organization
Division No. 4, Infantry
P.O. Box 2086
Fort Benning, GA 31905

1 Human Resources Research Organization
Division No. 5, Air Defense
P.O. Box 6057
Fort Bliss, TX 79916

1 Human Resources Research Organization
Division No. 6, Library
P.O. Box 428
Fort Rucker, AL 36360

1 Dr. Lawrence B. Johnson
Lawrence Johnson and Associates, Inc.
200 S Street, N.W., Suite 502
Washington, DC 20009

1 Dr. Norman J. Johnson
School of Urban and Public Affairs
Carnegie-Mellon University
Pittsburgh, PA 15213

1 Dr. David Klahr
Graduate School of Industrial
Administration
Carnegie-Mellon University
Pittsburgh, PA 15213

1 Dr. Robert R. Mackie
Human Factors Research, Inc.
6780 Cortona Drive
Santa Barbara Research Park
Goleta, CA 93017

1 Dr. Andrew R. Molnar
Technological Innovations in Education
National Science Foundation
Washington, DC 20550

1 Dr. Leo Munday, Vice President
American College Testing Program
P.O. Box 168
Iowa City, IA 52240

1 Dr. Donald A. Norman
Center for Human Information Processing
University of California, San Diego
La Jolla, CA 92037

1 Mr. Luigi Petrullo
2431 North Edgewood Street
Arlington, VA 22207

1 Dr. Robert D. Pritchard
Department of Psychology
Purdue University
Lafayette, IN 47907

1 Dr. Diane M. Ramsey-Klee
R-K Research & System Design
3947 Ridgemont Drive
Malibu, CA 90265

1 Dr. Joseph W. Rigney
Behavioral Technology Laboratories
University of Southern California
3717 South Grand
Los Angeles, CA 90007

1 Dr. Leonard L. Rosenbaum, Chairman
Department of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. George E. Rowland
Rowland and Company, Inc.
P.O. Box 61
Haddonfield, NJ 08033

1 Dr. Arthur I. Siegel
Applied Psychological Services
Science Center
404 East Lancaster Avenue
Wayne, PA 19087

1 Mr. Dennis J. Sullivan
725 Benson Way
Thousand Oaks, CA 91360

1 Dr. Benton J. Underwood
Northwestern University
Department of Psychology
Evanston, IL 60201

1 Dr. David J. Weiss
Department of Psychology
University of Minnesota
Minneapolis, MN 55455

1 Dr. Anita West
Denver Research Institute
University of Denver
Denver, CO 80210

1 Dr. Kenneth Wexler
School of Social Sciences
University of California
Irvine, CA 92664

1 Dr. John Annett
The Open University
Milton Keynes
Buckinghamshire, ENGLAND

1 Maj. P. J. DeLeo
Instructional Technology Branch
AF Human Resources Laboratory
Lowry AFB, CO 80230

1 Dr. Martin Rockway
Technical Training Division
Lowry Air Force Base
Denver, CO 80230

1 Dr. Eric McWilliams, Program Manager
Technology and Systems, TIE
National Science Foundation
Washington, DC 20550

1 Dr. Milton S. Katz
MITRE Corporation
Westgate Research Center
McLean, VA 22101

1 Dr. Charles A. Ullmann
Director, Behavioral Sciences Studies
Information Concepts Incorporated
1701 No. Ft. Myer Drive
Arlington, VA 22209

(Continued from inside front cover)

96 R. C. Atkinson, J. W. Brelsford, and R. M. Shiffrin. Multi-process models for memory with applications to a continuous presentation task. April 13, 1966. (J. math. Psychol., 1967, 4, 277-300).

97 P. Suppes and E. Crothers. Some remarks on stimulus-response theories of language learning. June 12, 1966.

98 R. Bjork. All-or-none subprocesses in the learning of complex sequences. (J. math. Psychol., 1968, 1, 182-195).

99 E. Gammon. The statistical determination of linguistic units. July 1, 1966.

100 P. Suppes, L. Hyman, and M. Jerman. Linear structural models for response and latency performance in arithmetic. (In J. P. Hill (ed.), Minnesota Symposia on Child Psychology. Minneapolis, Minn.: 1967. Pp. 160-200).

101 J. L. Young. Effects of intervals between reinforcements and test trials in paired-associate learning. August 1, 1966.

102 H. A. Wilson. An investigation of linguistic unit size in memory processes. August 3, 1966.

103 J. T. Townsend. Choice behavior in a cued-recognition task. August 8, 1966.

104 W. H. Batchelder. A mathematical analysis of multi-level verbal learning. August 9, 1966.

105 H. A. Taylor. The observing response in a cued psychophysical task. August 10, 1966.

106 R. A. Bjork. Learning and short-term retention of paired associates in relation to specific sequences of interpresentation intervals. August 11, 1966.

107 R. C. Atkinson and R. M. Shiffrin. Some Two-process models for memory. September 30, 1966.

108 P. Suppes and C. Ihrke. Accelerated program in elementary-school mathematics--the third year. January 30, 1967.

109 P. Suppes and I. Rosenthal-Hill. Concept formation by kindergarten children in a card-sorting task. February 27, 1967.

110 R. C. Atkinson and R. M. Shiffrin. Human memory: a proposed system and its control processes. March 21, 1967.

111 Theodore S. Rodgers. Linguistic considerations in the design of the Stanford computer-based curriculum in initial reading. June 1, 1967.

112 Jack M. Knutson. Spelling drills using a computer-assisted instructional system. June 30, 1967.

113 R. C. Atkinson. Instruction in initial reading under computer control: the Stanford Project. July 14, 1967.

114 J. W. Brelsford, Jr. and R. C. Atkinson. Recall of paired-associates as a function of overt and covert rehearsal procedures. July 21, 1967.

115 J. H. Stelzer. Some results concerning subjective probability structures with semiorders. August 1, 1967

116 D. E. Rumelhart. The effects of interpresentation intervals on performance in a continuous paired-associate task. August 11, 1967.

117 E. J. Fishman, L. Keller, and R. E. Atkinson. Massed vs. distributed practice in computerized spelling drills. August 18, 1967.

118 G. J. Groen. An investigation of some counting algorithms for simple addition problems. August 21, 1967.

119 H. A. Wilson and R. C. Atkinson. Computer-based instruction in initial reading: a progress report on the Stanford Project. August 25, 1967.

120 F. S. Roberts and P. Suppes. Some problems in the geometry of visual perception. August 31, 1967. (Synthese, 1967, 17, 173-201)

121 D. Jamison. Bayesian decisions under total and partial ignorance. D. Jamison and J. Kozielecki. Subjective probabilities under total uncertainty. September 4, 1967.

122 R. C. Atkinson. Computerized instruction and the learning process. September 15, 1967.

123 W. K. Estes. Outline of a theory of punishment. October 1, 1967.

124 T. S. Rodgers. Measuring vocabulary difficulty: An analysis of item variables in learning Russian-English and Japanese-English vocabulary pairs. December 18, 1967.

125 W. K. Estes. Reinforcement in human learning. December 20, 1967.

126 G. L. Wolford, D. L. Wessel, W. K. Estes. Further evidence concerning scanning and sampling assumptions of visual detection models. January 31, 1968.

127 R. C. Atkinson and R. M. Shiffrin. Some speculations on storage and retrieval processes in long-term memory. February 2, 1968.

128 John Holmgren. Visual detection with imperfect recognition. March 29, 1968.

129 Lucille B. Mlodnosky. The Frostig and the Bender Gestalt as predictors of reading achievement. April 12, 1968.

130 P. Suppes. Some theoretical models for mathematics learning. April 15, 1968 (Journal of Research and Development in Education, 1967, 1, 5-22)

131 G. M. Olson. Learning and retention in a continuous recognition task. May 15, 1968.

132 Ruth Norene Hartley. An investigation of list types and cues to facilitate initial reading vocabulary acquisition. May 29, 1968.

133 P. Suppes. Stimulus-response theory of finite automata. June 19, 1968.

134 N. Moler and P. Suppes. Quantifier-free axioms for constructive plane geometry. June 20, 1968. (In J. C. H. Gerretsen and F. Oort (Eds.), Compositio Mathematica. Vol. 20. Groningen, The Netherlands: Wolters-Noordhoff, 1968. Pp. 143-152.)

135 W. K. Estes and D. P. Horst. Latency as a function of number or response alternatives in paired-associate learning. July 1, 1968.

136 M. Schlag-Rey and P. Suppes. High-order dimensions in concept identification. July 2, 1968. (Psychom. Sci., 1968, 11, 141-142)

137 R. M. Shiffrin. Search and retrieval processes in long-term memory. August 15, 1968.

138 R. D. Freund, G. R. Loftus, and R.C. Atkinson. Applications of multiprocess models for memory to continuous recognition tasks. December 18, 1968.

139 R. C. Atkinson. Information delay in human learning. December 18, 1968.

140 R. C. Atkinson, J. E. Holmgren, and J. F. Juola. Processing time as influenced by the number of elements in the visual display. March 14, 1969.

141 P. Suppes, E. F. Loftus, and M. Jerman. Problem-solving on a computer-based teletype. March 25, 1969.

142 P. Suppes and Mona Morningstar. Evaluation of three computer-assisted instruction programs. May 2, 1969.

143 P. Suppes. On the problems of using mathematics in the development of the social sciences. May 12, 1969.

144 Z. Domotor. Probabilistic relational structures and their applications. May 14, 1969.

145 R. C. Atkinson and T. D. Wickens. Human memory and the concept of reinforcement. May 20, 1969.

146 R. J. Titiev. Some model-theoretic results in measurement theory. May 22, 1969.

147 P. Suppes. Measurement: Problems of theory and application. June 12, 1969.

148 P. Suppes and C. Ihrke. Accelerated program in elementary-school mathematics--the fourth year. August 7, 1969.

149 D. Rundus and R.C. Atkinson. Rehearsal in free recall: A procedure for direct observation. August 12, 1969.

150 P. Suppes and S. Feldman. Young children's comprehension of logical connectives. October 15, 1969.

(Continued on back cover)

(Continued from inside back cover)

151 Joaquim H. Laubsch. An adaptive teaching system for optimal item allocation. November 14, 1969.
152 Roberta L. Klatzky and Richard C. Atkinson. Memory scans based on alternative test stimulus representations. November 25, 1969.
153 John E. Holmgren. Response latency as an indicant of information processing in visual search tasks. March 16, 1970.
154 Patrick Suppes. Probabilistic grammars for natural languages. May 15, 1970.
155 E. Gammon. A syntactical analysis of some first-grade readers. June 22, 1970.
156 Kenneth N. Wexler. An automaton analysis of the learning of a miniature system of Japanese. July 24, 1970.
157 R. C. Atkinson and J.A. Paulson. An approach to the psychology of instruction. August 14, 1970.
158 R.C. Atkinson, J.D. Fletcher, H.C. Chetin, and C.M. Stauffer. Instruction in initial reading under computer control: the Stanford project. August 13, 1970.
159 Dewey J. Rundus. An analysis of rehearsal processes in free recall. August 21, 1970.
160 R.L. Klatzky, J.F. Juola, and R.C. Atkinson. Test stimulus representation and experimental context effects in memory scanning.
161 William A. Rottmayer. A formal theory of perception. November 13, 1970.
162 Elizabeth Jane Fishman Loftus. An analysis of the structural variables that determine problem-solving difficulty on a computer-based teletype. December 18, 1970.
163 Joseph A. Van Campen. Towards the automatic generation of programmed foreign-language instructional materials. January 11, 1971.
164 Jamesine Friend and R.C. Atkinson. Computer-assisted instruction in programming: AID. January 25, 1971.
165 Lawrence James Hubert. A formal model for the perceptual processing of geometric configurations. February 19, 1971.
166 J. F. Juola, I.S. Fischler, C.T. Wood, and R.C. Atkinson. Recognition time for information stored in long-term memory.
167 R.L. Klatzky and R.C. Atkinson. Specialization of the cerebral hemispheres in scanning for information in short-term memory.
168 J.D. Fletcher and R.C. Atkinson. An evaluation of the Stanford CAI program in initial reading (grades K through 3) March 12, 1971.
169 James F. Juola and R.C. Atkinson. Memory scanning for words versus categories.
170 Ira S. Fischler and James F. Juola. Effects of repeated tests on recognition time for information in long-term memory.
171 Patrick Suppes. Semantics of context-free fragments of natural languages. March 30, 1971.
172 Jamesine Friend. Instruct coders' manual. May 1, 1971.
173 R.C. Atkinson and R.M. Shiffrin. The control processes of short-term memory. April 19, 1971.
174 Patrick Suppes. Computer-assisted instruction at Stanford. May 19, 1971.
175 D. Jamison, J.D. Fletcher, P. Suppes, and R.C. Atkinson. Cost and performance of computer-assisted instruction for compensatory education.
176 Joseph Offir. Some mathematical models of individual differences in learning and performance. June 28, 1971.
177 Richard C. Atkinson and James F. Juola. Factors influencing speed and accuracy of word recognition. August 12, 1971.
178 P. Suppes, A. Goldberg, G. Kanz, B. Searle, and C. Stauffer. Teacher's handbook for CAI courses. September 1, 1971.
179 Adele Goldberg. A generalized instructional system for elementary mathematical logic. October 11, 1971.
180 Max Jerman. Instruction in problem solving and an analysis of structural variables that contribute to problem-solving difficulty. November 12, 1971.
181 Patrick Suppes. On the grammar and model-theoretic semantics of children's noun phrases. November 29, 1971.
182 Georg Kreisel. Five notes on the application of proof theory to computer science. December 10, 1971.
183 James Michael Moloney. An investigation of college student performance on a logic curriculum in a computer-assisted instruction setting. January 28, 1972.
184 J.E. Friend, J.D. Fletcher, and R.C. Atkinson. Student performance in computer-assisted instruction in programming. May 10, 1972.
185 Robert Lawrence Smith, Jr. The syntax and semantics of ERICA. June 14, 1972.
186 Adele Goldberg and Patrick Suppes. A computer-assisted instruction program for exercises on finding axioms. June 23, 1972.
187 Richard C. Atkinson. Ingredients for a theory of instruction. June 26, 1972.
188 John D. Bonvillian and Veda R. Charrow. Psycholinguistic implications of deafness: A review. July 14, 1972.
189 Phipps Arabie and Scott A. Boorman. Multidimensional scaling of measures of distance between partitions. July 26, 1972.
190 John Ball and Dean Jamison. Computer-assisted instruction for dispersed populations: System cost models. September 15, 1972.
191 W. R. Sanders and J. R. Ball. Logic documentation standard for the Institute for Mathematical Studies in the Social Sciences. October 4 ,1972.
192 M.T. Kane. Variability in the proof behavior of college students in a CAI course in logic as a function of problem characteristics. October 6, 1972.
193 P. Suppes. Facts and fantasies of education. October 18, 1972.
194 R. C. Atkinson and J. F. Juola. Search and decision processes in recognition memory. October 27, 1972.
195 P. Suppes, R. Smith , and M. Léveillé. The French syntax and semantics of PHILIPPE, part 1 : Noun phrases. November 3, 1972.
196 D. Jamison, P. Suppes, and S. Wells. The effectiveness of alternative instructional methods: A survey. November 1972.
197 P. Suppes. A survey of cognition in handicapped children. December 29, 1972.
198 B. Searle, P. Lorton, Jr., A. Goldberg, P. Suppes, N. Ledet, and C. Jones. Computer-assisted instruction program: Tennessee State University. February 14, 1973.
199 D. R. Levine. Computer-based analytic grading for German grammar instruction. March 16, 1973.
200 P. Suppes, J.D. Fletcher, M. Zanotti, P. V. Lorton, Jr., and B. W. Searle. Evaluation of computer-assisted instruction in elementary mathematics for hearing-impaired students. March 17, 1973.
201 G. A. Huff. Geometry and formal linguistics. April 27, 1973.
202 C. Jensema. Useful techniques for applying latent trait mental-test theory. May 9, 1973.
203 A. Goldberg. Computer-assisted instruction: The application of theorem-proving to adaptive response analysis. May 25, 1973.
204 R. C. Atkinson, D. J. Herrmann, and K. T. Wescourt. Search processes in recognition memory. June 8, 1973.
205 J. Van Campen. A computer-based introduction to the morphology of Old Church Slavonic. June 18,1973.
206 R. B. Kimball. Self-optimizing computer-assisted tutoring: Theory and practice. June 25, 1973.
207 R.C. Atkinson, J.D. Fletcher, E. J. Lindsay, J.O. Campbell, and A. Barr. Computer-assisted instruction in initial reading. July 9, 1973.
208 V. R. Charrow and J. D. Fletcher. English as the second language of deaf students. July 20, 1973.
209 J. A. Paulson. An evaluation of instructional strategies in a simple learning situation. July 30, 1973.